THE UNIVERSITY
of ADELAIDE

Faculty of Sciences
School of Physical Sciences

Extending the Energy Range of Ultra-High Energy Cosmic Ray Fluorescence Detectors

Justin Mark Thomas-Albury

Submitted for the degree of
Doctor of Philosophy

Supervisors:
Dr. José A Bellido
Prof. Bruce R Dawson

November 2020

Justin Mark Thomas-Albury

Extending the Energy Range of Ultra-High Energy Cosmic Ray Fluorescence Detectors

November 2020

Supervisors: Dr. José A Bellido and Prof. Bruce R Dawson

The University of Adelaide
School of Physical Sciences
Faculty of Sciences
North Terrace
Adelaide 5005

Contents

Abstract

The origin, acceleration mechanisms and mass composition of ultra-high energy cosmic rays (UHECR) are among the biggest unsolved mysteries in astrophysics. The search is limited by the extremely low number of cosmic rays observed at the highest energies; 1 particle per square kilometre per year. At these energies, significant numbers of cosmic rays cannot be measured directly. Instead, extensive air showers (EAS) produced by their interaction with the atmosphere are measured by ground-based experiments. One such experiment, the Pierre Auger Observatory (Auger), adopts a hybrid detection technique utilising fluorescence telescopes and surface detector stations to measure the energy, arrival direction and chemical mass composition of UHECRs.

The Pierre Auger Observatory (Auger) measures the depth of shower maximum (X_{max}), an observable which is sensitive to the mass composition of the primary cosmic ray. A modified analysis chain has been developed for the study of the chemical mass composition of UHECRs which extends the energy range of existing analysis to both lower and higher energies. This has been achieved by relaxing strict data selection cuts, while carefully accounting for the detector efficiency and acceptance using Monte Carlo (MC) simulations. Furthermore, the systematic uncertainties associated with the new analysis technique are studied and discussed in detail. The findings of this analysis are compared with previous results from the Pierre Auger Observatory (Auger) as well as other experiments.

A novel technique for the reconstruction of EAS using Fluorescence Detector (FD) measurements has been developed in this thesis. The Fluorescence detector Array of Single-pixel Telescopes (FAST) is a design concept for the next generation of UHECR detectors which will target UHECRs above $10^{19.5}$ eV with unprecedented sensitivity. The reconstruction algorithm has been developed in two parts: a machine learning approach, and a top-down maximum likelihood technique. The top-down reconstruction is based on the comparison of measured data with the output of a sophisticated detector simulation. In a proof-of-concept study, it has been shown that the Fluorescence detector Array of Single-pixel Telescopes (FAST) will be capable of reconstructing EAS with acceptable resolution. The results of this thesis pave the way for FAST to become a next-generation UHECR detector.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

November 2020

–
Justin Mark Thomas-Albury

Acknowledgements

I would like to begin by thanking my supervisors, Dr Jose Bellido and Professor Bruce Dawson, whose expertise, support and advice has been invaluable during my PhD. Without their guidance and constant feedback, this would not have been possible. I have thoroughly enjoyed my time working with them over the last few years.

I would also like to thank the High Energy Astrophysics research group at the University of Adelaide. In particular I would like to thank Gary Hill and Roger Clay for their advice and input to my research. I would also like to express special thanks to Violet Harvey and Jarryd Day for all of their help over our years of study together.

Thank you to all members of the Pierre Auger Collaboration and the FAST Collaboration, without which, this research would not be possible. In particular I would like to thank Toshihiro, Dusan, Max and John who I have worked closely with during my PhD. I appreciate all of their help and support.

The support of my family has been crucial in me accomplishing such a great achievement. In particular, I want to thank my parents, Lyn and Andrew, and my sister Paige. They have always encouraged me to follow my passions, and I am extremely grateful for their unfailing support and encouragement. Finally, I want to thank Olivia for her unwavering love, support and encouragement.

1 Introduction

The origin and acceleration mechanisms of ultra-high energy cosmic rays (UHECR) have yet to be elucidated. Measurement of the energy spectrum and mass composition of the highest energy cosmic rays is essential to understanding the origin, acceleration mechanisms and possible sources of the highest energy cosmic rays. Such measurements also constrain fundamental astrophysical and particle physics models. Current generation UHECR observatories like the Pierre Auger Observatory (Auger) measure the energy spectrum and mass composition of cosmic rays in unprecedented detail through the detection and reconstruction of extensive air showers (EAS). However, Auger is only sensitive to a particular energy range of the spectrum. Auger is limited at lower energy by sensitivity and trigger efficiency, and at high energy by the extremely low flux of the highest energy cosmic rays.

This thesis will focus on extending the energy range of UHECR measurements, with a particular focus on fluorescence detectors. The first part of this thesis is focused on extending the energy range of the Auger mass composition analysis, and the second part explores a new detector design optimised for the measurement of UHECRs above $10^{19.5}$ eV.

Mass Composition Studies at the Pierre Auger Observatory

The Fluorescence Detectors (FDs) of Auger measure the depth of shower maximum X_{max}, which is correlated with the mass composition of the primary cosmic ray particle. The standard mass composition analysis employs important data selection techniques to account for inefficiencies in the detector which would otherwise cause significant biases in X_{max} measurements. This is extremely important for accurate comparisons with hadronic interaction models.

The X_{max} analysis has previously been limited to energies above $10^{17.2}$ eV due to the inefficiency of the Surface Detector (SD) trigger algorithms and problems with the data section algorithms not producing expected results. However, with the addition of new SD trigger algorithms, the X_{max} analysis can be extended to lower energies without composition bias if a new method of accounting for the detector acceptance can be developed and an accurate estimate of the SD trigger probabilities can be determined. This extension to lower energy is extremely important for comparisons with other experiments which operate in a lower energy range. In addition to the

lower energy extension, a higher energy extension of the X_{max} measurements above $10^{19.7}$ eV will provide important insight to the origin and acceleration mechanisms of cosmic rays at the highest energies.

The Fluorescence detector Array of Single-pixel Telescopes (FAST)

Existing detectors can be upgraded, and analysis techniques can be extended to improve measurements, however, it has become clear that a much larger area must be instrumented in order to measure properties of the energy spectrum and mass composition of UHECRs with sufficient statistical significance. The Fluorescence detector Array of Single-pixel Telescopes (FAST) is a prototype detector which could address these requirements. Unfortunately, the low-cost design of the FAST FD telescopes comes at the cost of reduced low energy performance. Even at the highest energies, standard reconstruction techniques are not viable for FAST since the telescopes contain only four pixels in the same field of view (FoV) of standard Auger telescopes which contain 440 pixels.

In the second part of this thesis, a novel technique for the reconstruction of EAS measured by FAST is developed and thoroughly tested. New reconstruction techniques must be developed if FAST is to become a viable next-generation UHECR detector. The author has developed all reconstruction software, including a full event data structure. Previous studies on possible reconstruction algorithms for FAST have only considered the case where a companion SD array provides an estimate of the shower geometry, leaving only X_{max} and energy for the FAST telescopes to determine. The extension of the FAST reconstruction to include the shower geometry is essential for the viable operation of a future FAST array.

The structure and content of each chapter is summarised below:

Chapter 2: Ultra-High Energy Cosmic Rays

An introduction to cosmic rays including a brief overview of the history and their discovery. The most recent measurement results are discussed including the energy spectrum, origins and mass composition of UHECRs. The physics of extensive air showers and how they are used to measure the properties of ultra-high energy cosmic rays is also discussed. Finally, an overview of previous and current cosmic ray experiments is presented.

Chapter 3: The Pierre Auger Observatory

An overview of the Pierre Auger Observatory including descriptions of the observatory detectors, trigger, calibration and reconstruction algorithms is presented in this chapter.

Chapter 4: Energy Threshold for Full Trigger Efficiency

A study of the trigger probability and trigger efficiency of the Surface Detectors of the Pierre Auger Observatory. The results of this chapter are used in Chapter 5 to extend the energy range of Auger's mass composition analysis.

Chapter 5: Mass Composition Studies

A study investigating the possibility of extending the energy range of the current X_{max} analysis for the Pierre Auger Observatory, using the trigger efficiency results from Chapter 4, and a new technique for estimating the detector X_{max} acceptance. A new analysis method is developed, and the systematic uncertainties of the method are studied in detail. The final results of the analysis are compared to previous results from the Pierre Auger Observatory, as well as to the results of other experiments.

Chapter 6: The Fluorescence Detector Array of Single-Pixel Telescopes

A brief introduction to the Fluorescence detector Array of Single pixel Telescopes (FAST) including motivations for the experiment, prototype detector descriptions and recent results.

Chapter 7: Top-Down Reconstruction

Introduction of a novel technique for the reconstruction of extensive air showers with the Fluorescence detector Array of Single-pixel Telescopes (FAST) based on a top-down algorithm which relies on a sophisticated detector simulation. The top-down reconstruction algorithm is developed and tested on simulated events in this chapter. The top-down reconstruction technique is shown to be viable in two different modes; a hybrid mode, and a full array mode. However, it is sensitive to the first guess of the shower parameters. Therefore, following this chapter, a first guess reconstruction algorithm is developed in Chapter 8.

Chapter 8: Neural Network Reconstruction

An alternate approach to the reconstruction of extensive air showers measured by the Fluorescence detector Array of Single-pixel Telescopes (FAST) using machine learning techniques. A neural network is trained with simulated showers in this chapter in order to develop a model which can be used to provide first guesses of the shower parameters to the top-down reconstruction developed in Chapter 7. The performance of this reconstruction technique is studied in detail, including the systematic uncertainties associated with the neural network approach.

Chapter 9: FAST Reconstruction Performance

The reconstruction performance of the Fluorescence detector Array of Single-pixel Telescopes (FAST) using the top-down reconstruction in its two modes is studied

in detail. The results of Chapters 7 and 8 provide a complete reconstruction chain for FAST including a first guess and final estimate of important shower parameters. Additionally, case studies of the reconstruction of real measured events measured by FAST prototype telescopes are presented.

2 Ultra-High Energy Cosmic Rays

Ultra-high energy cosmic rays (UHECR) are the highest energy particles in the universe, with observed energies up to 10^{20} eV. They consist primarily of protons and other atomic nuclei. Cosmic rays provide information about the dynamics of the highest energy astrophysical events and environments in the universe, as well as information about the interstellar medium, including the strength and structure of the magnetic fields through which they propagate. Understanding the origins, acceleration mechanisms, propagation and mass composition of the highest energy cosmic rays are some the key challenges in astroparticle physics.

2.1 A Brief History of Cosmic Ray Physics

In 1912, Austrian physicist Victor Hess discovered cosmic rays. His discovery was based on the work of many other scientists, including French physicist Henri Becquerel, who discovered the natural radioactivity of certain chemical elements in 1896 [1], and Marie and Pierre Curie, who discovered the radioactivity of Radium. The electroscope, a device that was sensitive to ionisation produced by radioactive elements, was one of the primary research instruments used to study radioactivity at the time. French physicist Charles-Augustin de Coulomb was the first to discover an unusual spontaneous discharge in electroscopes at the end of the 18$^{\text{th}}$ century [2]. This was also observed independently by Michael Faraday in 1835 [3]. At the beginning of the 20$^{\text{th}}$ century, it was believed that the cause of the spontaneous electroscope discharge was the ionisation of the air inside the device induced by terrestrial radioactivity.

In 1901, Scottish physicist Charles Thomson Rees Wilson performed numerous electroscope experiments and was the first to suggest that the ionisation could be caused by radiation from beyond the Earth's atmosphere. He took his equipment underground to test this hypothesis, hoping to detect a reduction in the ionisation rate due to absorption by the rock above. Unfortunately, he was unable to achieve any conclusive results [4]. Theodor Wulf attempted to test the hypothesis that the ionisation of the air inside the electroscope was the result of terrestrial radiation. Wulf carried a portable electroscope to the top of the Eiffel Tower. He expected the observed ionisation rate to decrease at higher altitude. This would provide ev-

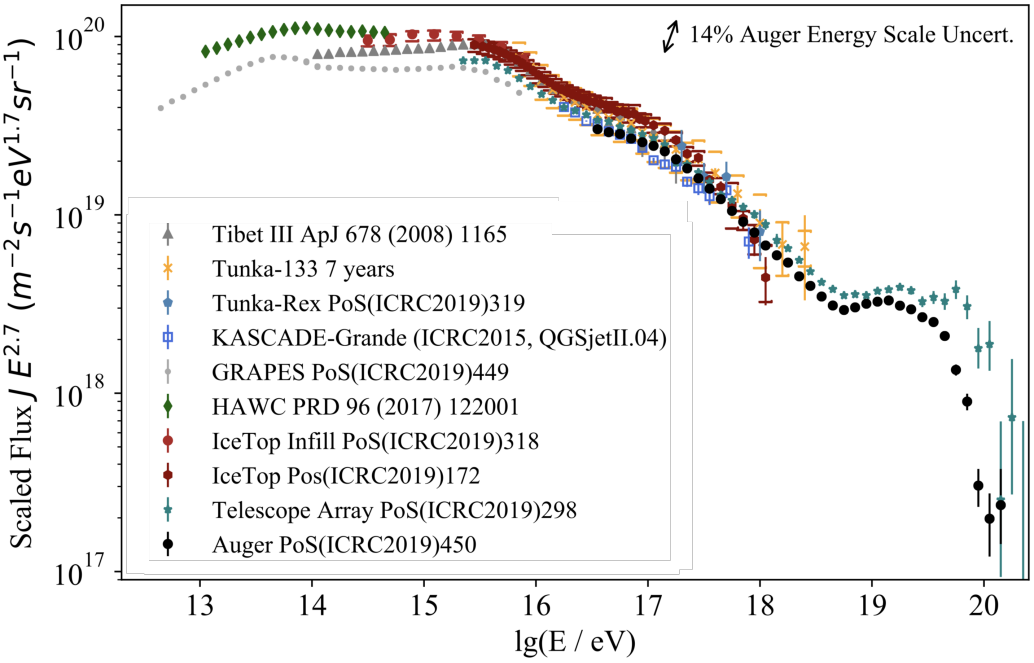


Figure 2.1. The cosmic ray energy spectrum as measured by a number of experiments. The flux is scaled by $E^{2.7}$ to clearly show the changes in the slope [13].

idence that the radiation was of terrestrial origin. However, his results were also inconclusive [5]. In 1910, Italian physicist Domenico Pacini tested ionisation rates at different altitudes, including underwater. He was able to show a decrease in the ionisation rate below the surface of the water, which contradicted the assumption that terrestrial radiation was the cause. However, it was Victor Hess who later performed similar measurements of ionisation in a hot air balloon, reaching altitudes of up to 5400 m. He was able to conclusively show that, although the radiation initially decreased at altitude, when he reached an altitude of around 2000 m, the radiation began to increase significantly [6]. This led to the surprising conclusion that a component of the radiation originated in outer space. Hess received the 1936 Nobel Prize in Physics for his discovery [7].

2.2 Energy Spectrum

The observed differential energy spectrum in Figure 2.1 shows the arrival rate of UHECRs at Earth as a function of energy. The spectrum follows an inverse power law $E^{-\gamma}$

$$\frac{dN}{dE} \propto E^{-\gamma}\ \mathrm{m^2 s^{-1} sr^{-1} eV^{-1}} \tag{2.1}$$

which extends over many orders of magnitude in energy [8–12]. The spectral index, γ is almost constant at ~2.7. There are, however, a number of interesting features in the spectrum where the spectral index differs [14]. These variations in

the spectral index are thought to correspond to transitions between different cosmic ray sources.

2.2.1 The Knees

The first of these features is a steepening known as the "knee" at an energy of $\sim 3 \times 10^{15}$ eV where the spectral index changes from 2.7 to ~3.1 [12, 15, 16]. The knee is generally considered to be a reduction in flux due to the fact that the sources of acceleration within our own galaxy are limited to a few times 10^{15} eV. It may also be linked to the fact that the proton gyro-radius is larger than that of the heavier nuclei (see Section 2.3.3). This means that protons are capable of escaping the galaxy above energies corresponding to the knee. The decrease in flux is then due to the higher energy protons that escape the galaxy more easily than heavier nuclei, and thus never reach Earth [17, 18].

There is an additional steepening of the spectrum at the "second knee" at an energy of $\sim 10^{17}$ eV [12]. Results from the KASCADE experiment suggest that the second knee corresponds to a hardening in the spectra of heavy elements, and a softening in the spectra of light elements [19, 20]. With recent improvements in analysis techniques and increased statistics, the Pierre Auger Observatory (Auger) has been able to measure the second knee to a high degree of precision, showing that the feature is a softening of the spectrum rather than a sharp change. It extends over an energy range of 100-200 PeV [21, 22]. The origin of the second knee is not clear; however, it may be attributed to the maximum Galactic acceleration energy of heavy nuclei as explained by a rigidity-dependent cut-off model [15, 23]. It is also possible that the lowest energy protons from extra-galactic sources contribute to the flux in the region of the second knee [23–25].

2.2.2 The Ankle

There is also a change of flux at the so called "ankle" which can be seen clearly in Figures 2.1 and 2.2. This change is observed as a flattening of the spectrum at $\sim 5 \times 10^{18}$ eV [22, 26, 27] and is considered to be the transition from galactic to extra-galactic sources where the flux of incoming extra-galactic cosmic rays begins to dominate the spectrum. It is likely that only at these high energies will the flux of galactic cosmic rays be reduced enough to observe an appreciable extra-galactic flux [8]. This is discussed further in Section 2.3.4.

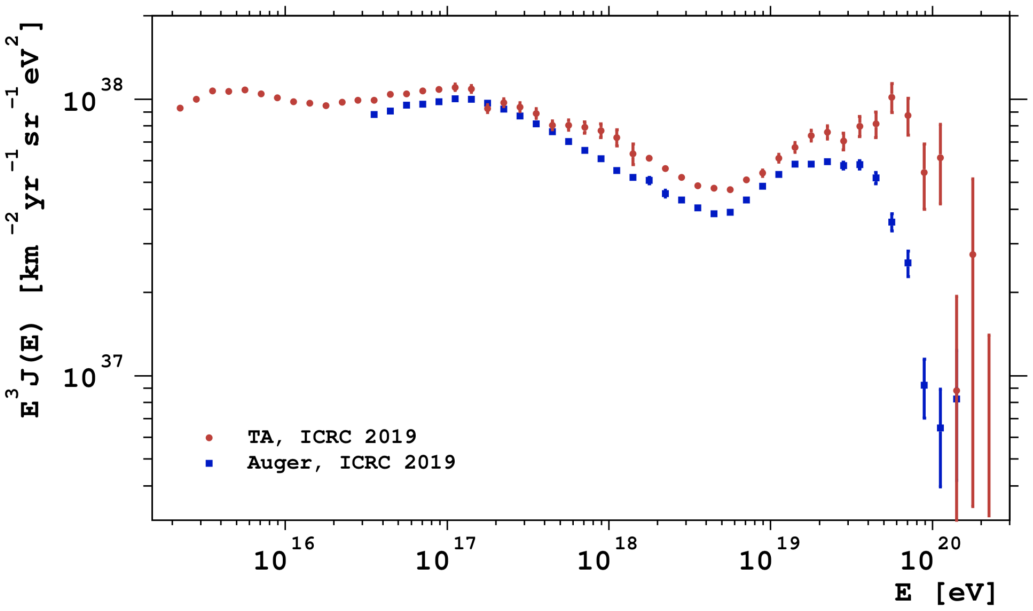


Figure 2.2. The cosmic-ray energy spectrum measured by the Pierre Auger Observatory (Auger) and the Telescope Array (TA) [28].

2.2.3 The Suppression

Finally, a rapid reduction in the flux at $\sim 6 \times 10^{19}$ eV has been observed. It was first measured with statistical significance by the High Resolution Fly's Eye (HiRes) experiment using fluorescence detector measurements [29, 30]. This suppression has been observed in both the Northern and Southern Hemispheres at the Telescope Array (TA) and Auger, respectively[1] [12, 28]. There are, however, still some unexplained differences between the spectra of the two experiments (shown in Figure 2.2), even in the common declination band observed by both observatories [28]. Recently due to Auger's increased precision, it has been discovered that the suppression is showing additional structure. It appears that an additional feature between the ankle and the suppression is required to explain the shape of the spectrum [22].

The suppression can be interpreted as either the maximum acceleration energy of the sources or a cut-off by an effect known as the Greisen-Zatsepin-Kuz'min (GZK) limit. The GZK limit arises due to the loss of energy in collisions between UHECRs and cosmic microwave background (CMB) photons [31–33]. Protons with energies above 6×10^{19} eV interact with the CMB photons and rapidly lose energy due to photo-pion production

$$p + \gamma_{\text{CMB}} \rightarrow n + \pi^{+} \tag{2.2}$$

$$p + \gamma_{\text{CMB}} \rightarrow p + \pi^{0} \tag{2.3}$$

[1] The Pierre Auger Observatory is located in Mendoza, Argentina in the Southern Hemisphere and the Telescope Array is located in Utah, USA in the Northern Hemisphere.

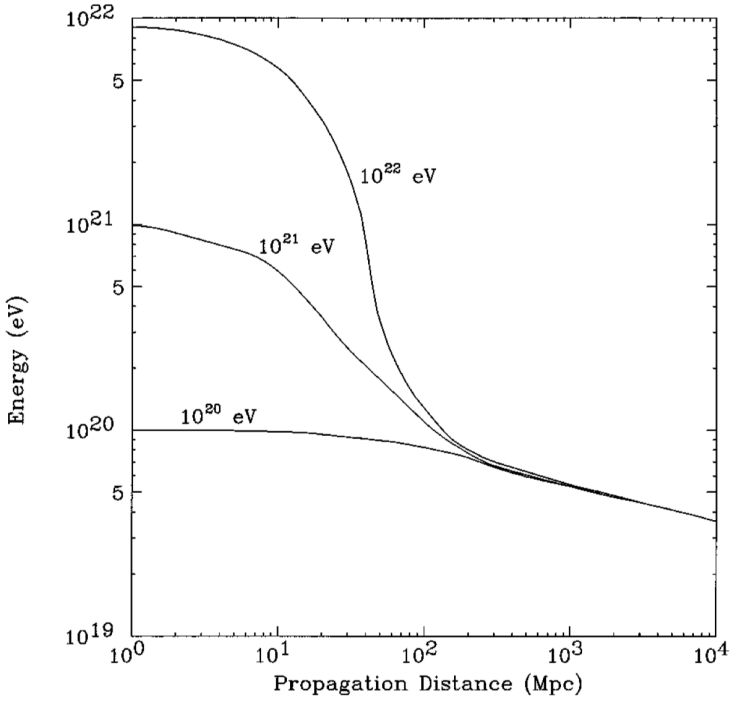


Figure 2.3. The energy of cosmic ray protons versus propagation distance. Energy losses are due to interactions with CMB photons. Notice that even if cosmic rays are accelerated to higher energies, then they quickly lose energy and converge to this limit within about 100 Mpc [36].

where neutral or charged pions are produced. In the case of the production of a charged pion, the proton is converted to a neutron. Heavier nuclei suffer from photo-disintegration and pair production [34, 35]. The implications of this energy loss are that cosmic rays with energies above the GZK limit would never reach the Earth unless the source of their acceleration was very close. In order to measure cosmic rays with energies greater than 10^{20} eV, the source of their acceleration would have to be located within approximately 100 Mpc of Earth, as depicted in Figure 2.3. Had such cosmic rays originated from within the Milky Way, there would have been a strong anisotropy of cosmic rays. The lack of such a strong anisotropy indicates that these UHECRs are of extra-galactic origin. In fact, Auger sees an anisotropy in the opposite direction to the Galactic centre [37, 38]. This is discussed further in Section 2.3.4.

The shape of the cosmic ray energy spectrum presents a significant challenge to the detection of UHECRs due to the low flux of cosmic rays at the ultra-high energies:

- Above 10^{11} eV: 1 particle per square metre per second
- Above the knee: 1 particle per square metre per year

- Above the ankle: 1 particle per square kilometre per year
- Above the suppression: 1 particle per square kilometre per century.

Therefore, due to limited statistics, the direct detection of UHECRs using satellites or balloons is not feasible. An indirect measurement of cosmic rays through the study of extensive air showers (EAS) using a very large ground-based detector is the solution for increasing statistics. This concept will be discussed in further detail in Section 2.4.

2.3 Origin of Cosmic Rays

The study of the origin of the highest energy cosmic rays is based on a number of fundamental questions: What are their sources? By what mechanisms do they reach such high energies? And how do they propagate through the interstellar medium (ISM) to arrive at Earth? The answers to these questions are still largely unknown, however significant progress has been made in recent years towards a deeper understanding.

2.3.1 Potential Sources

There are two subsets of UHECR source models: *top-down* and *bottom-up* models. Top-down scenarios propose that the observed energetic particles originate from the decay of super-massive particles possibly from the early universe. Bottom-up models, on the other hand, suggest that cosmic rays are accelerated in extreme astrophysical environments and gain energy until they eventually escape.

2.3.1.1 Top-down models

Top down (non-acceleration) scenarios for the origin of cosmic rays easily avoid problems in explaining the mechanisms by which UHECRs are accelerated to energies of the order of 10^{20} eV. Instead, the particles' origins lie in the decay of some unknown super-massive particles of mass $\geq 10^{20}$ eV. Their decay products therefore have energies of up to 10^{20} eV. Furthermore, these super-massive particles do not actually have to be associated with any specific astrophysical source [39]. Top-down scenarios predict the decay into UHECRs to result in the production a large flux of photons, among other particles including electrons, positrons and neutrinos. For this reason, ultra-high energy photon and neutrino limits are powerful tools to discriminate between top-down and bottom-up scenarios [40, 41]. Experimental

limits on the flux of high-energy photons strongly disfavour top-down scenarios [42–44].

2.3.1.2 Bottom-up models

Bottom-up models suggest that UHECRs are the result of astrophysical shock acceleration of charged particles. Naturally, the Fermi acceleration models produce a power-law spectrum similar to the spectrum discussed in Section 2.2. Such models are attractive; however, these acceleration models must be able to explain $\geq 10^{20}$eV energy particles. This scenario will be discussed further in Section 2.3.2.

2.3.2 Acceleration Mechanisms

Although the mechanisms by which cosmic rays are accelerated to the highest energies are still unknown, several models have been developed to explain their enormous energies. A notable contribution comes from Enrico Fermi in 1949 [45], who first suggested a stochastic acceleration model, now known as Fermi acceleration. Fermi's model was later extended to describe cosmic ray interactions with astrophysical shocks, in a model known as diffusive shock acceleration (DSA) [46].

2.3.2.1 Fermi Acceleration

Fermi's original acceleration model, called second-order Fermi acceleration, is a stochastic model which considers the interactions between cosmic ray particles and gas clouds in the ISM. These gas clouds have essentially random velocities $\sim$ 15 km/s [47]. Consider an ISM cloud travelling at velocity v and a charged particle (cosmic ray) with initial energy E_1 and momentum p_1 which enters the cloud at angle θ_1 with respect to the cloud's trajectory. In the cloud reference frame, the initial energy of the particle is given by

$$E_1' = \gamma E_1 \left(1 - \beta \cos\theta_1\right) \tag{2.4}$$

where $\beta = v/c$ and $\gamma = 1/\sqrt{1-\beta^2}$. The particle scatters elastically and emerges in a random direction with no change in energy ($E_2' = E_1'$). This process is depicted in Figure 2.4. Transforming Equation (2.4) to an inertial reference frame, the particle's energy after collision becomes

$$E_2 = \gamma E_2' \left(1 - \beta \cos\theta_2'\right) \tag{2.5}$$

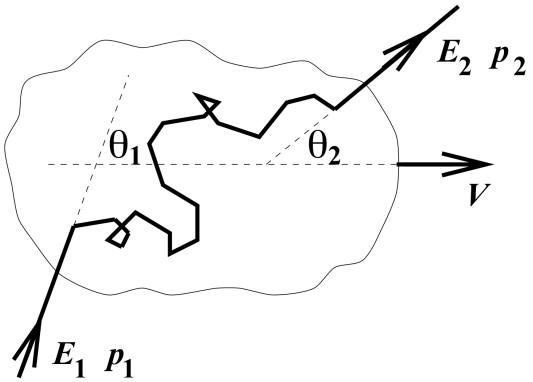


Figure 2.4. A diagram of the interaction of a cosmic ray particle of energy E_1 with an ISM cloud moving at velocity v [47].

The fractional energy change in the inertial frame can then be defined as

$$\frac{\Delta E}{E} = \frac{E_2 - E_1}{E_1} = \frac{1 - \beta\cos\theta_1 + \beta\cos\theta_2' - \beta^2\cos\theta_1\cos\theta_2'}{1-\beta^2} - 1 \tag{2.6}$$

The average fractional energy change can be determined based on the average angles of collision and scattering, $\cos\theta_1$ and $\cos\theta_2'$. Since the particle scatters such that its direction is random upon emerging from the cloud, the average value of $\cos\theta_2'$ can simply be written as

$$\langle\cos\theta_2'\rangle = 0 \tag{2.7}$$

The average value of $\cos\theta_1$, however, is dependent on the rate of collisions between the particle and cloud, which is proportional to their relative velocities. The average value of $\cos\theta_1$ is given by

$$\langle\cos\theta_1\rangle = -\frac{\beta}{3} \tag{2.8}$$

Substituting Equations (2.7) and (2.8) into Equation (2.6) yields

$$\frac{\langle\Delta E\rangle}{E} = \frac{1+\beta^2/3}{1-\beta^2} - 1 \approx \frac{4}{3}\beta^2 \tag{2.9}$$

So, it is clear that the fractional energy change is positive and proportional to β^2. However, this provides only a small energy gain since $\beta \ll 1$, due to the fact that the probability of a head-on collision and energy gain is only slightly higher than that of a tail-on collision where energy is lost [47].

2.3.2.2 Diffusive Shock Acceleration

Fermi's original theory was extended to a so called first-order Fermi acceleration model in order to improve the efficiency of energy gain in the model [46, 48–50]. The new DSA model considers the many repeated interactions of cosmic ray particles with ISM gas clouds, taking place in strong astrophysical shocks such as supernovae.

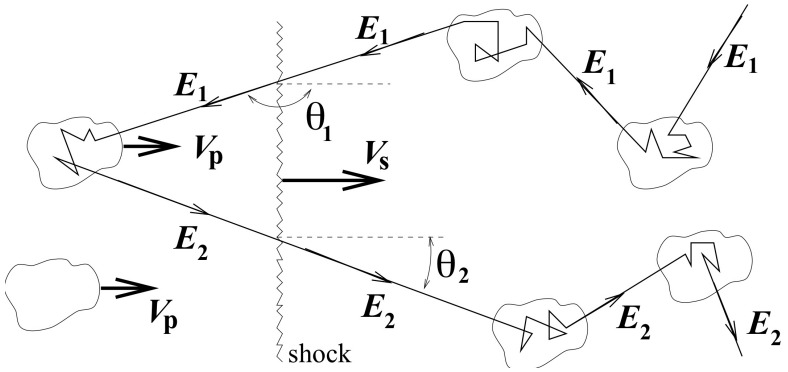


Figure 2.5. A diagram of the interaction of a cosmic ray particle of energy E_1 with a shock front moving at velocity v_s [47].

During a supernova, several solar masses of material can be ejected with speeds of $\sim 10^4$ km/s, significantly breaking the "sound" barrier of ~10 m/s in the ISM. An associated shock front propagates radially outwards, causing the ISM material and associated magnetic fields to pile up at the shock front. The velocity of the shock, v_s depends heavily on the velocity of the ejecta, v_p and the compression ratio, R, according to the following equation

$$\frac{v_s}{v_p} = \frac{R}{R-1} \tag{2.10}$$

Shock hydrodynamics theory shows that strong supernova shocks have $R = 4$ [47].

As the shock front propagates through the ISM, the cosmic ray particle has some probability of colliding and interacting with an ISM cloud. After this interaction, there is a chance the particle will be ejected back towards the shock front where the process repeats, and the particle continues to gain energy as it bounces back and forth across the shock as depicted in Figure 2.5. In a similar way to second-order Fermi acceleration, the average fractional energy gain of this process can be estimated using the average interaction angles $\langle \cos\theta_1 \rangle$ and $\langle \cos\theta_2' \rangle$.

$$\langle \cos\theta_1 \rangle = -\frac{2}{3} \tag{2.11}$$

$$\langle \cos\theta_2' \rangle = \frac{2}{3} \tag{2.12}$$

Therefore, the average fractional energy gain is given by

$$\frac{\langle \Delta E \rangle}{E} \approx \frac{4}{3}\beta \tag{2.13}$$

providing a significantly more efficient acceleration mechanism. That can be explained by the fact that whatever side of the shock the particle is on, it sees gas

clouds approaching at velocity v_p on the other side of the shock, which greatly increases the likelihood of head-on collisions.

2.3.2.3 Energy Spectrum

The resulting differential energy spectrum given the DSA model can be derived by considering the number of cosmic rays in the shock, and the rate at which they escape from the shock. Consider the number of particles that escape the shock after k crossings given an initial number of particles N_0,

$$N_k = N_0 P(1-P)^{k-1} \tag{2.14}$$

that is, the number of particles which escape at the kth step is the number which have survived the previous $k-1$ steps, multiplied by the escape probability P. Since the probability of returning to the shock is given by $1-P$, the probability of crossing the shock and returning k times is given by

$$P(\text{cross} \geq k) = (1-P)^k \,. \tag{2.15}$$

The energy after k crossings is therefore

$$E_k = E_0 \left(1 + \frac{\Delta E}{E}\right)^k \,. \tag{2.16}$$

The number of particles at energy E can then be determined by

$$\frac{dN}{dE} = N_k \frac{dE(k)}{dk} = N_k E_k \ln(1+f) \tag{2.17}$$

where $f = \Delta E/E$, leading to the form

$$\frac{dN}{dE} = \frac{N_0 P}{(1-P)E_0 \ln(1+f)} \left(\frac{E_k}{E_0}\right)^{\frac{\ln(1-P)}{\ln(1+f)}-1} \tag{2.18}$$

So we have the differential energy spectrum which is of the same form as Equation (2.1), with the spectral index

$$\gamma = \frac{\ln(1-P)}{\ln(1+f)} - 1 \approx \frac{R+2}{R-1} \tag{2.19}$$

So, for $R = 4$, the energy spectrum becomes E^{-2} compared to the observed spectrum below the knee of $E^{-2.7}$. The observed spectrum is likely steepened because the probability of escape from the shock is larger than the fractional energy gain, that is $P > f$. This could be explained by the energy dependent escape of cosmic rays from the Galaxy [47].

2.3.2.4 Possible Acceleration Regions

Given the model of DSA, the maximum acceleration energy of an acceleration source can be derived based on its size. In fact, the energy of a source is constrained by the Larmor radius or gyro-radius of a particle as it passes through a magnetic field of strength B. The Larmor radius is given by

$$r_{\mathrm{L}} = \frac{p_{\perp}}{ZeB} \tag{2.20}$$

where $p_{\perp}$ is the particle's momentum perpendicular to the magnetic field and Ze is the particle's charge [51]. As the Larmor radius of a particle approaches the size of the acceleration region, it becomes increasingly likely that the particle will escape. This general dimensionality limit is referred to as the Hillas criterion and imposes a maximum acceleration energy

$$E_{\mathrm{max}} \leq \gamma ZeBr_{\mathrm{L}} \tag{2.21}$$

where γ is the Lorentz factor of the shock. This relation is reflected in the Hillas plot [52] shown in Figure 2.6. The Hillas plot shows the possible accelerators capable of accelerating cosmic rays to the highest energies. Possible accelerators of UHECRs are gamma-ray burst (GRB), active galactic nucleis (AGNs), supernova remnants (SNRs), galaxy clusters, starburst galaxies, giant radio galaxies, pulsars and intergalactic medium (IGM) shocks [53]. The Hillas criterion is a necessary condition for DSA, but not a sufficient one. Hillas [52] discusses other source constraints, including those imposed by energy losses in photon fields.

2.3.3 Propagation

The trajectories of cosmic ray particles are influenced by the magnetic fields which exist within the Galaxy and in extragalactic space. The galactic magnetic field has a regular component of ~3-4 μG (3×10^{-10} T) which is associated with the spiral arms of the galaxy [51, 54]. A reasonable (but not certain) estimate of the extragalactic magnetic field strength is $\sim 10^{-9}$ G [55]. These magnetic fields cause deflection of charged particles by the Lorentz force with a radius of gyration (or Larmor radius) given by

$$r = \frac{p}{ZeB} = \frac{\rho}{cB} \tag{2.22}$$

where p is the momentum of the particle, Ze is its charge, B is the magnetic field and ρ is the magnetic rigidity

$$\rho = \frac{E}{ZecB}. \tag{2.23}$$

This means that the deflection of a cosmic ray particle depends on both its energy and its charge. Protons with energy $\sim 10^{20}$ eV in the Galactic magnetic field, would

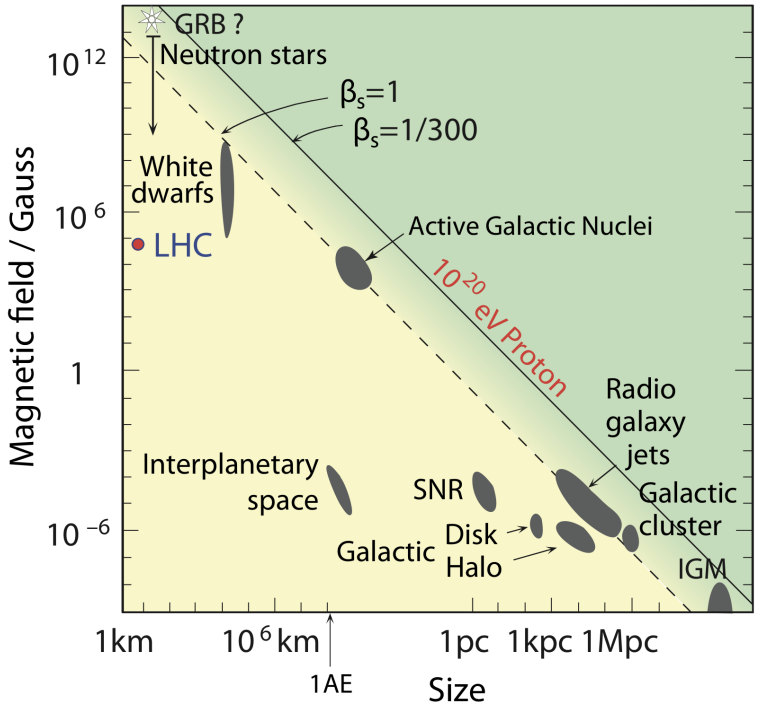


Figure 2.6. A Hillas plot [52] representing the potential astrophysical acceleration sources for cosmic rays based on the Hillas criterion for maximum energy.

have a gyro-radius on the order of 100 kpc (significantly larger than the ∼16 kpc radius of the Galaxy). This suggests 10^{20} eV protons accelerated within the Galaxy would travel on almost straight trajectories. The lack of a clear source within our own galaxy suggests that the highest energy cosmic rays must originate from beyond our own galaxy.

Extragalactic magnetic fields are not well understood but are expected to produce random deflections in cosmic ray trajectories [56]. There are also smaller scale irregular components of the Galactic magnetic field which combine with the regular component to give a total Galactic magnetic field strength of ∼6 μG [54]. An energetic charged particle will scatter off of these irregularities (in the Galactic or extragalactic magnetic fields) on the scale of its gyro-radius, resulting in diffusive propagation. If the particle's energy is large enough (or mass small enough), these irregularities are negligible, however lower energy (or more massive) particles, with smaller gyro-radii, can become trapped in the magnetic field lines of the irregular component, resulting in the randomisation of their trajectory through a random-walk process [56].

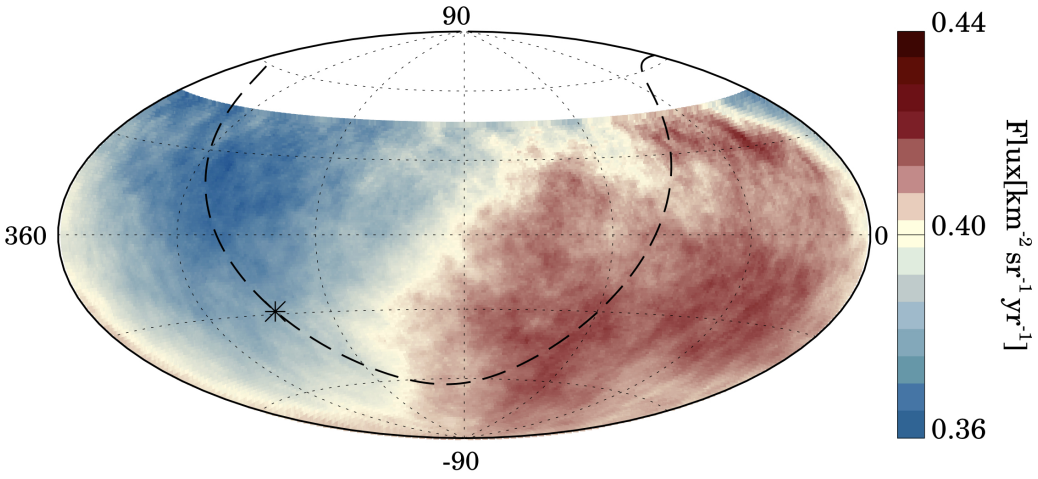


Figure 2.7. A map in equatorial coordinates of the cosmic ray flux above 8 EeV from the Pierre Auger Observatory (Auger). The Galactic centre is indicated as a star. The flux is averaged within a top-hat window of 45° in radius. A clear excess flux in a direction nearly opposite to that of the Galactic centre can be seen [37].

2.3.4 Arrival Directions

Since charged particles follow curved and randomised trajectories through magnetic fields, cosmic rays tend to arrive almost isotropically at Earth. This means that their arrival directions do not usually indicate their true origin. Only at the highest energies, where these deviations are not as large, do their trajectories follow a more direct path from their source. Therefore, it is difficult to determine the origin of their acceleration. Nevertheless, anisotropy studies are powerful tools for the search of the origin of cosmic rays.

The Pierre Auger Observatory (Auger) has measured a large-scale anisotropy at the highest energies. A dipole modulation with an amplitude of $6.6^{+1.2}_{-0.8}\,\%$ and pointing $\sim 125^\circ$ away from the Galactic centre is observed above 8 EeV [37, 38, 57]. A map of the cosmic ray flux and the dipole amplitude and phase as a function of energy are shown in Figures 2.7 and 2.8, respectively. These results from Auger indicate an extra-galactic origin for cosmic rays above 8 EeV, while lower energy results (below $\sim$1 EeV) indicate a dipole in the direction of the Galactic centre [58–60]. Additional evidence for the extra-galactic origin of UHECRs has been found by Auger with the correlation of cosmic ray arrival directions with nearby starburst galaxies. The maximum likelihood is found for starburst galaxies at energies above 38 EeV with a post-trial significance of 4.5 σ [61, 62]. No conclusive evidence for such a correlation with starburst galaxies is found in the Telescope Array (TA) data [63].

Evidence for two intermediate scale anisotropies has been observed; the hotspot in the Northern Hemisphere observed by TA [64, 65], and the hotspot around

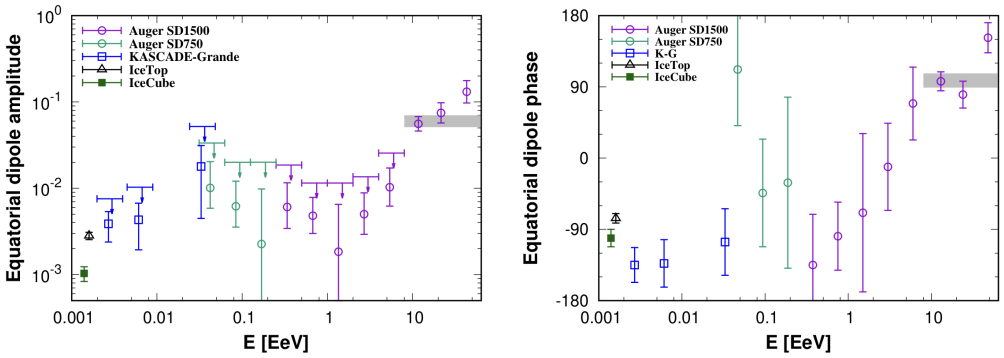


Figure 2.8. Amplitude (left) and phase (right) as a function of energy of the Auger dipole [37]. Lower energy results from IceCube, IceTop and KASCADE-Grande are also shown [58–60].

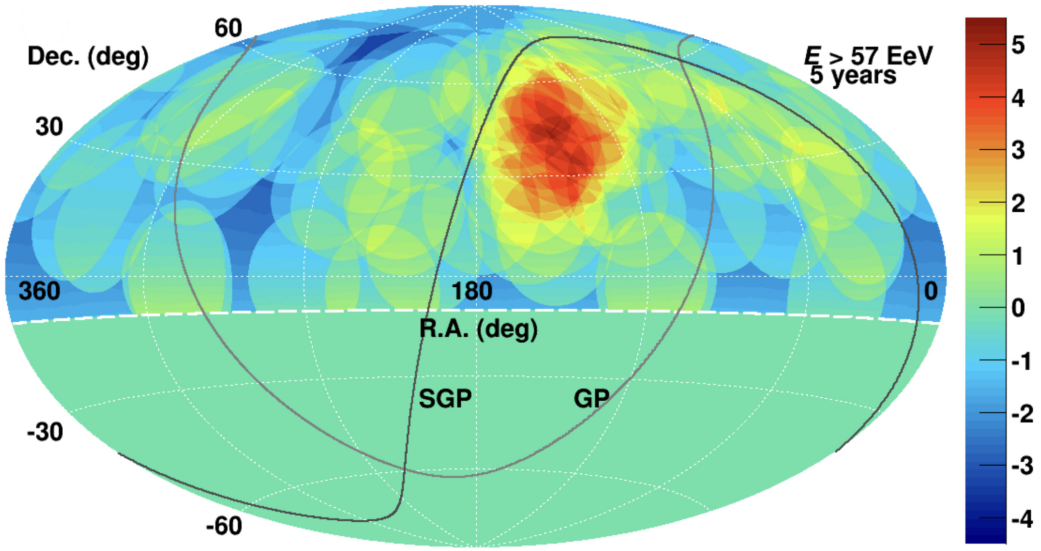


Figure 2.9. Significance map of the TA hotspot with energy above 57 EeV from events observed between May 2008 and May 2013 [64].

Centaurus A observed by Auger in the Southern Hemisphere. The TA hotspot, shown in Figure 2.9, is observed above 57 EeV and is located in the direction of the Ursa Major cluster. While the first five years of TA data suggested a post-trial significance of 3.4 σ [64], after 11 years of data, that was reduced to 2.9 σ [65]. The Centaurus A excess observed by Auger rejects the isotropic hypothesis with a post-trial significance of 3.9 σ [61, 66]. Neither of these hotspots from TA or Auger are statistically significant enough to reach the discovery level. More recently a joint anisotropy search using data from both TA and Auger has found evidence for a possible quadrupolar pattern in the observed large-scale anisotropy [67].

2.4 Extensive Air Showers

The cosmic ray energy spectrum, as discussed in Section 2.2, follows a steep inverse power law, with a decrease in the flux of cosmic rays with energy over many orders of magnitude. Above $\sim 10^{14}$ eV, the extremely low flux of cosmic rays inhibits their direct detection. Instead, very large ground-based detectors are used to measure the phenomenon known as an extensive air shower (EAS). When a cosmic ray (or another particle such as a photon) first interacts with the Earth's atmosphere, it undergoes electromagnetic or nuclear interactions with atmospheric molecules such as nitrogen or oxygen. These interactions form a cascade of secondary particles, each subsequently interacting with atmospheric molecules and producing additional secondary particles. EAS were discovered experimentally by French physicist, Pierre Auger in the 1930's using Geiger counters and Wilson (cloud) chambers. He observed particles arriving in time coincidence at different detectors spaced up to 300 m apart, eventually leading to the conclusion that they were in fact secondary particles originating from a common primary particle [68].

Interactions between a cosmic ray such as a proton and an atmospheric nucleus produce many secondary particles including pions and kaons. A diagram of the shower components is shown in Figure 2.10. An EAS can be separated into three components:

1. **Electromagnetic component**
   Neutral pions are produced in the interactions between cosmic rays and atmospheric nuclei and subsequently decay into gamma rays, producing the electromagnetic component of the shower, which consists of electrons, positrons and gamma rays.

2. **Hadronic component**
   The hadronic component consists of protons, neutrons, charged pions and kaons. High energy charged pions can also interact with atmospheric nuclei to produce additional pions.

3. **Muonic component**
   The decay of lower energy charged pions and kaons produce muons and neutrinos which form the muonic component of the shower. High energy muons will be detected at the ground, however low energy muons can again decay into electrons and positrons before reaching the ground.

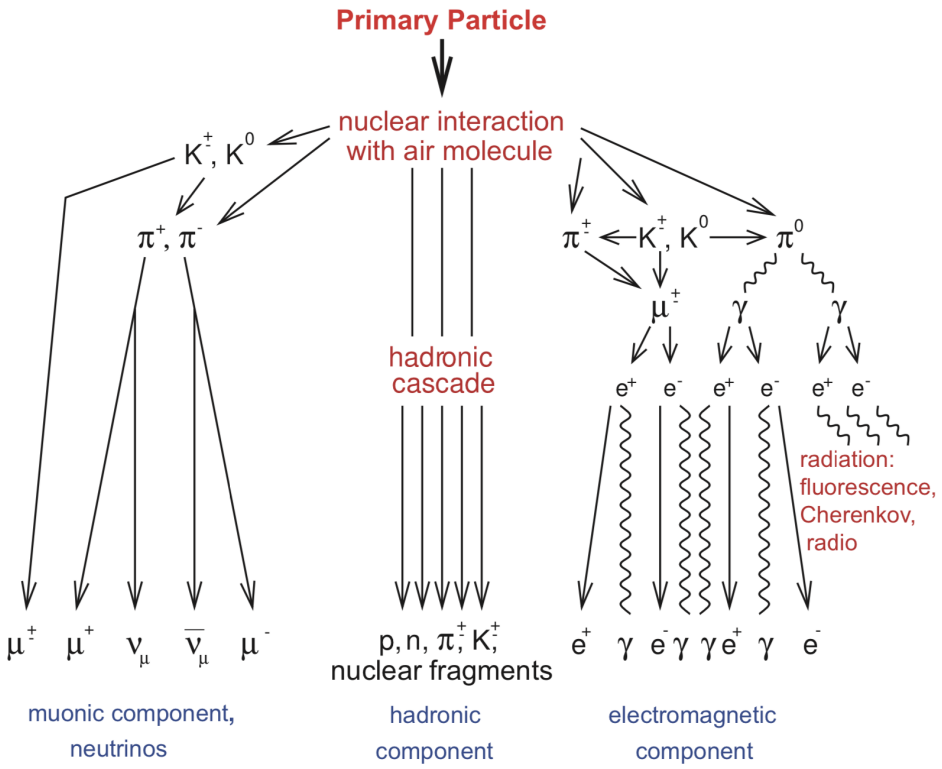


Figure 2.10. A diagram showing the components of an extensive air shower (EAS) initiated by a hadronic primary cosmic ray [69].

2.4.1 Electromagnetic Cascades

Electromagnetic cascades are a series of electromagnetic interactions of photons and electrons with atmospheric nuclei. These cascades are initiated by either high energy photons or electron-position pairs produced by the decay of pions in the hadronic component of an EAS. Pions decay into gamma-rays or electron-position pairs via the following decay chains

$$\pi^0 \to \gamma + \gamma \qquad (\sim 98.8\%) \tag{2.24a}$$

$$\pi^0 \to \gamma + e^+ + e^- \qquad (\sim 1.2\%) \tag{2.24b}$$

Alternatively, they can be initiated directly by a high energy primary photon. In the case of a photon as the primary particle, soon after the photon enters the atmosphere, it interacts with an atmospheric nucleus and undergoes pair production. This results in the production of an electron-positron pair which share the energy of the initial photon. The electrons and positrons undergo Bremsstrahlung which causes photon emission during the interaction with the electromagnetic fields of atmospheric nuclei. In either case, the energy of the electron or photon is equally divided amongst the decay products.

Heitler proposed a model for the description of electromagnetic air showers [70]. The model provides a first-order approximation of the shower development based on the assumption that photons, electrons and positrons undergo simple two-body interactions after a fixed radiation length, producing new particles. A schematic diagram of an electromagnetic cascade is shown in Figure 2.11(left). Initially, a photon with energy E_0 interacts with an atmospheric nucleus (N) resulting in the production of a positron-electron pair

$$\gamma + N \rightarrow N + e^+ + e^- \tag{2.25}$$

with the electron and positron each (on average) carrying away half the energy of the incident photon, $E_0/2$. At an atmospheric depth

$$X = \lambda_r \ln 2 \tag{2.26}$$

where λ_r is the radiation length in the medium ($\sim$37 g/cm^2 for air [71]). The electron-positron pair lose energy after travelling an additional depth of X and produce a gamma ray with half their energy via Bremsstrahlung

$$e^{\pm} + N \rightarrow e^{\pm} + N + \gamma \tag{2.27}$$

In reality, the energy is likely to be divided between several photons, so the model overestimates the number of particles. The gamma rays produced by Bremsstrahlung will continue to produce additional electron-positron pairs through pair production. The process continues in this way until the energy of the particles approaches the critical energy[2], E_c, the energy at which collisional energy losses dominate over radiative energy losses. Once the maximum number of particles is reached, the shower particles gradually lose energy through ionisation and the emission of soft photons and the shower eventually dies out.

The depth of shower maximum is obtained based on the number of interaction lengths, n, required for the energy per particle to be reduced to critical energy, E_c [72]. Based on initial photon energy E_0, after n interaction lengths (a distance of $nX = n\lambda_r \ln 2$), the total number of particles in the shower is $N = 2^n$. The energy of each shower particle will be

$$E(n) = \frac{E_0}{N} \tag{2.28}$$

The shower reaches its maximum size, N_{max} when most particles have energy E_c, that is

$$N_{\text{max}} = \frac{E_0}{E_c} \tag{2.29}$$

[2] $E_c = 85$ MeV for air [72].

Combining Equations (2.28) and (2.29) yields

$$E(n_{\text{max}}) = E_c = \frac{E_0}{2^{n_{\text{max}}}} \tag{2.30}$$

which leads to the number of radiation lengths before the shower reaches its maximum development which is given by

$$n_{\text{max}} = \frac{\ln\left(E_0/E_c\right)}{\ln 2} \tag{2.31}$$

The depth of shower maximum, or X^{γ}_{max} for an electromagnetic shower, is therefore given by

$$X^{\gamma}_{\text{max}} = n_{\text{max}} \times X \tag{2.32a}$$

$$= \lambda_r \ln\left(E_0/E_c\right) \tag{2.32b}$$

The elongation rate, Λ, is defined as the rate of change in X_{max} per decade of primary energy, E_0

$$\Lambda = \frac{dX_{\text{max}}}{d\log_{10} E_0} \tag{2.33}$$

combining this with Equation (2.32) yields

$$\Lambda^{\gamma} = 2.3\lambda_r = 85 \text{ g/cm}^2/\text{dec} \tag{2.34}$$

The Heitler model underestimates the ratio of photons to leptons due to several of the simplifying assumptions. This is partly due to Bremsstrahlung interactions occasionally producing more than one photon. Furthermore, the model overestimates the shower maximum due to the simplified modeling of energy loss mechanisms. Although the Heitler model is simple, it correctly predicts shower development up to X_{max}, in particular that $X_{\text{max}} \propto \ln E_0$ and $N_{\text{max}} \propto E_0$ [72]. It will be seen in Chapter 3 that this approximation is reasonable, however experiment shows that the energy deposited does not fall to zero quickly, but instead has a long tail.

2.4.2 Hadronic Cascades

In the case where the primary particle is an atomic nucleus, the EAS cascade instead begins with hadronic interactions with atmospheric nuclei. The hadronic component of an EAS involves many multi-particle interactions as well as various decay processes which makes it significantly more difficult to model compared to the electromagnetic component. The atmospheric depth at which the first interaction occurs depends strongly on the mass of the primary particle [73]. For lighter nuclei like protons, the first interaction occurs deeper in the atmosphere at approximately 70 g/cm^2 at energies around the spectral knee. On the other hand, for heavier

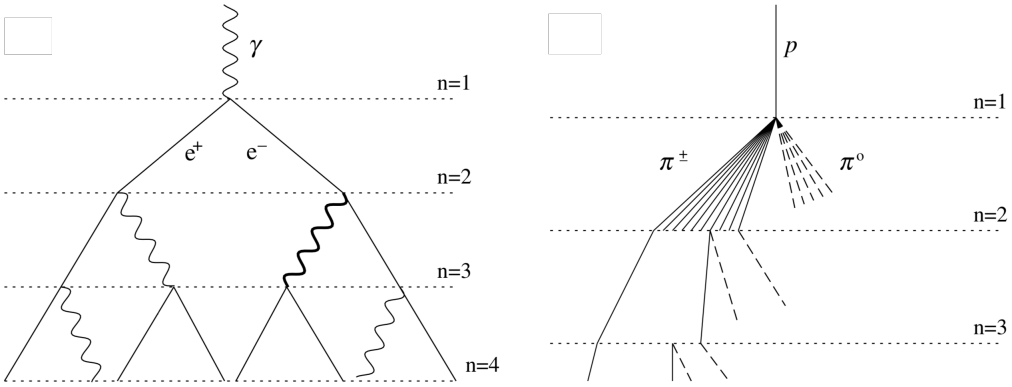


Figure 2.11. Schematic views of (left) an electromagnetic cascade with electrons and positrons indicated by solid lines and gamma-rays indicated by waves, and (right) a hadronic cascade with solid lines indicating charged pions and dashed lines indicating neutral pions which quickly decay to produce electromagnetic sub-showers [72].

nuclei like iron the first interaction occurs at a shallower atmospheric depth of approximately 15 g/cm^2. This is equivalent to saying that the proton-air cross section for interaction is significantly less compared to the iron-air cross section. These hadronic interactions produce pions and kaons of which approximately one third are neutral pions [72]. As discussed in Section 2.4.1, neutral pions decay into gamma rays or electron-position pairs which initiate the electromagnetic component of a shower. The majority of pions produced are charged. These charged pions either interact again, or decay into muons and neutrinos which form the muonic component of the shower.

$$\pi^+ \rightarrow \mu^+ + \nu_\mu \tag{2.35a}$$

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu \tag{2.35b}$$

The majority of muons produced in the shower reach the ground since time dilation extends their lifetime in the reference frame of the ground. The muon content of a shower is useful to discriminate between hadronic showers and electromagnetic showers. Some of the resulting muons decay into electrons or positrons (depending on charge). These products also contribute to the electromagnetic component of the shower [72].

$$\mu^+ \rightarrow e^+ + \bar{\nu}_\mu + \nu_e \tag{2.36a}$$

$$\mu^- \rightarrow e^- + \nu_\mu + \bar{\nu}_e \tag{2.36b}$$

To understand the development of a hadronic shower through the atmosphere, Matthews extended the Heitler model of an electromagnetic cascade to the case of a shower initiated by a proton or heavier nucleus [72]. The resulting model

considers the population of pions and their interactions over the course of many interactions lengths[3], $X = \lambda_I \ln 2$. Consider a hadron primary with initial energy E_0. Since the neutral pions (one third of the population) are lost to the electromagnetic component of the shower via their decay into gamma rays, after n interactions the energy remaining in the hadronic cascade is given by

$$E = \left(\frac{2}{3}\right)^n E_0 \tag{2.37}$$

and assuming equal division between flavours of pions, the total number of charged pions will be

$$N_\pi = (N_{\text{ch}})^n \tag{2.38}$$

where N_{ch} is multiplicity of charged particles produced in hadronic interactions[4] . The energy per charged pion after n interactions will therefore become

$$E_\pi = \frac{E_0}{N_\pi}\left(\frac{2}{3}\right)^n = \frac{E_0}{\left(\frac{3}{2}N_{\text{ch}}\right)^n} \tag{2.39}$$

The charged pions continue on to the next interaction where the process repeats until the energies of the charged pions fall below the critical energy E_c^π, at which point they are all assumed to decay into muons which reach the ground [72]. A schematic diagram of a hadronic cascade is shown in Figure 2.11(right). As for the electromagnetic cascade, the number of interaction lengths required for the energy per particle to be reduced to the critical energy is the number of interactions required to obtain $E_\pi = E_c$. Following from Equation (2.39) the number of interactions to reach E_c becomes

$$n_{\text{max}} = \frac{\ln\left(E_0/E_c\right)}{\ln\left(\frac{3}{2}N_{\text{ch}}\right)} \tag{2.40}$$

The depth of shower maximum is determined by considering the first generation of neutral pions produced in the hadronic cascade. The photons produced as their decay products initiate electromagnetic sub-showers each of energy $E_0/\left(3N_{\text{ch}}\right)$ which develop in parallel. The depth of shower maximum can be estimated by the sum of the first interaction depth (the depth at which the pions are first produced) and the depth of maximum of the electromagnetic sub-showers generated by the pion decay. Combining this with Equation (2.32) yields

$$X^p_{\text{max}} = X_0 + \lambda_r \ln\left(\frac{E_0}{3N_{\text{ch}}E_c}\right) \tag{2.41}$$

The model underestimates X_{max} compared to detailed EAS simulations, which is likely a consequence of ignoring subsequent generations of pions which also generate

[3]For pions in air, $\lambda_I \approx 120$ g/cm^2
[4]Matthews adopts a constant value of $N_{\text{ch}} = 10$ in his model.

electromagnetic sub-showers [72]. The elongation rate can also be derived as for the electromagnetic cascade. It is useful to rewrite Equation (2.41) in terms of the depth of maximum for the electromagnetic cascade from Equation (2.32) as

$$X^{p}_{\text{max}} = X^{\gamma}_{\text{max}} + X_0 - \lambda_r \ln\left(3N_{\text{ch}}\right) \tag{2.42}$$

The elongation rate for proton showers is then

$$\Lambda^{p} = \Lambda^{\gamma} + \frac{d}{d\log_{10} E_0}\left(X_0 - \lambda_r \ln\left(3N_{\text{ch}}\right)\right) \tag{2.43}$$

$$= 58 \text{ g/cm}^2\text{/dec} \tag{2.44}$$

The model can be extended to larger mass primaries using the superposition principle, where a shower initiated by a hadron of mass A is approximated as A proton sub-showers of energy E_0/A. It follows from Equation (2.32) and Equation (2.41) (which both conclude $X_{\text{max}} \propto \ln E_0$) that

$$X^{A}_{\text{max}} \propto \ln \frac{E_0}{A} \tag{2.45}$$

from which it can be concluded that showers initiated by lighter nuclei penetrate deeper into the atmosphere before reaching their maximum compared to showers initiated by heavier nuclei.

2.4.3 Measurement of Extensive Air Showers

The measurement of EAS is a method of indirect detection of cosmic rays. By studying various observables from ground level, information about the primary cosmic ray can be inferred. The most well established and widely used detection methods are ground arrays which measure the lateral distribution of particles at the ground, and air fluorescence telescopes (for UHECRs) which measure the longitudinal development of the shower as it traverses the atmosphere. At lower energies (around the spectral knee), smaller ground arrays (e.g. KASCADE) and Cherenkov detectors are favoured. Cherenkov detectors measure the faint Cherenkov light emitted by relativistic shower particles as the propagate through the atmosphere.

2.4.3.1 Ground Arrays

The lateral distribution of the signal from an EAS can be measured at ground level using a sparse array of surface detectors. The benefit of ground arrays is their 100% duty cycle. This is a significant advantage for the statistically limited analysis of UHECRs. Unfortunately, ground arrays do not observe the development of the shower

directly but must instead reconstruct information about the primary particle from the particles which are measured at the ground. The lateral distribution of signals can be used as an estimate of the primary cosmic ray's energy with a typical resolution of $20 - 30\%$ [14]. The relative trigger timing can be used to determine its arrival direction with a resolution of better than 1° [14]. Some common ground-based detectors are:

- **Water-Cherenkov detectors** - consist of a sealed tank filled with purified water which emits Cherenkov light from relativistic shower particles passing through the water. The emitted Cherenkov light is measured by photomultiplier tubes (PMTs) which are placed on the inside of the tank. This type of detector is discussed further in Chapter 3.

- **Scintillation detectors** - consist of a scintillating material which absorbs energy from incoming charged particles and re-emits a portion of it in the form of light pulses. These light pulses can then be measured by PMTs within the detector which convert the light into an electronic signal. Scintillators can be shielded by a material which partially absorbs the electromagnetic component of the shower, thereby allowing the direct measurement of the muonic component.

- **Radio detectors** - consist of radio antennas which can an be used to measure the radio and micro-waves emitted by air showers. There are two main effects causing radio emission of air showers. The dominant mechanism arises from geomagnetically induced transverse currents caused by the Lorentz force acting on electrons and positrons in the shower. These transverse currents grow as the shower develops, reaching a maximum and then declining with the electromagnetic component of the shower. A secondary mechanism arises from a negative charge excess (known as the Askaryan effect) where ionised atmospheric electrons propagate with the shower, leaving the more massive positively charged ions behind. This charge excess grows and declines with the shower development. It is the time variation of the transverse currents and the charge excess which causes radio emission of air showers [74].

2.4.3.2 Air Fluorescence Telescopes

Unlike surface detectors, fluorescence telescopes can only operate with a duty cycle of around 15-20% since they require dark skies (sun and moon below the horizon) and good weather conditions. The fluorescence approach relies on measuring the faint fluorescence light emitted isotropically in the wavelength range of 300-400 nm by excited molecular nitrogen in the atmosphere. This makes it feasible to directly

image the development of an EAS as it traverses the atmosphere from a side-on view and at large distances. This allows for an almost calorimetric measure of the energy of the primary particle, with only the energy from muons and neutrinos unaccounted for. Fluorescence telescopes are discussed further in Chapters 3 and 6.

2.5 Mass Composition

The mass composition of cosmic rays is a one of the essential observables which can be used to understand the features of the energy spectrum discussed in Section 2.2. Below 10^{14} eV direct measurements of the composition of cosmic rays is possible using high-altitude balloon measurements and satellite-based detectors. However, at the highest energies, measurements of the mass composition of cosmic rays is only possible via observations of EAS. For Fluorescence Detector (FD) measurements, observations of the depth of shower maximum, X_{max}, is especially well suited for composition studies. While other observables are also suitable, X_{max} is the convenient because the required measurement resolution is well within the capabilities of FDs. For a given primary energy, X_{max} is dependent on the mass composition of the primary particle which initiated the air shower; as such, it can be used to discriminate particles by their primary mass. Typically, the X_{max} distributions are compared to predictions of various hadronic models to infer the average mass composition as a function of energy. On the other hand, for Surface Detector (SD) measurements, estimates of mass composition can be made based on the relative abundance of muons and electromagnetic particles in the shower at ground level [75], or by other methods (such azimuthal asymmetry or rise time) that rely on development of EAS.

Event-by-event discrimination of mass composition is not possible at the highest energies due to fluctuations in EAS development. However, the average mass composition can be inferred from shower observables which correlate with the mass composition of the primary particle. As discussed in Section 2.4, X_{max} is inversely proportional to the logarithm of the atomic mass of the primary particle. This implies that at a given energy, showers with higher primary masses are initiated at shallower atmospheric depths than those with lighter masses. Furthermore, the assumption that a shower initiated by a primary of mass A can be approximated as A proton sub-showers (Section 2.4.2) leads to the conclusion that shower-to-shower fluctuations decrease in magnitude with increasing mass. Therefore, the width of the X_{max} distribution is expected to decrease with increasing mass, making it another powerful mass composition indicator. Based on interpretations of the X_{max} distribution, its mean, $\langle X_{\text{max}} \rangle$ and its standard deviation, $\sigma(X_{\text{max}})$ conclusions can be made about the average mass composition of cosmic rays as a function of

energy. The mean of the X_{max} distribution alone is not sensitive to composition dispersion, since a given measurement of $\langle X_{\text{max}} \rangle$ can be attributed to different mass composition mixtures. For example, a mixture of proton and iron can produce the same $\langle X_{\text{max}} \rangle$ as a mixture of helium and nitrogen. On the other hand, $\sigma(X_{\text{max}})$ is sensitive to the separation in the components of the X_{max} distribution. In the previous example, $\sigma(X_{\text{max}})$ would distinguish the two different mass composition mixtures. To interpret $\langle X_{\text{max}} \rangle$ and $\sigma(X_{\text{max}})$ comparisons with high energy hadronic interaction models need to be made.

2.5.1 Hadronic Interaction Models

The particle interactions which take place in an EAS occur in a phase-space well beyond the reach of particle accelerators. The centre-of-mass energy of 14 TeV achieved by the Large Hadron Collider (LHC) corresponds to an energy of 10^{17} eV in the laboratory frame, while UHECRs reach up 10^{20} eV (centre-of-mass energy of 450 TeV) [76]. High energy interaction models are used to extrapolate the measurements from accelerators to the UHECR region. The latest hadronic interaction models include QGSJetII-04 [77, 78], EPOS-LHC [79], and Sibyll 2.3 [80] and Sibyll 2.3c [81]. Each of these models are unique and offer different scenarios for EAS measurements, the differences of their predictions of X_{max} for proton and iron primaries are shown in Figure 2.12. The depth of shower maximum of an individual shower is highly dependent of the depth of first interaction. Therefore, the proton-air cross-section is the most direct connection between EAS development and hadronic interactions. A larger proton-air cross-section (shorter interaction length) results in a shallower depth of first interaction. This results in a shallower X_{max} and hence smaller $\sigma(X_{\text{max}})$. As shown in Figure 2.12, Sibyll 2.3c predicts the average shower maximum for protons to be $\sim$20 g/cm^2 deeper compared to Sibyll 2.1. This difference is largely attributed to a reduction in the proton-air cross-section in the updated model [81]. Additionally, a decreased inelasticity for proton and pion interactions also has the effect of increasing X_{max} and $\sigma(X_{\text{max}})$ since particles retain a larger fraction of their energy during interactions.

The predictions of hadronic models are extremely important for the interpretation of cosmic ray mass composition. As such, important cross validations of the models have been done. This has led to the discovery of a clear muon deficit in the predictions of hadronic models compared to that of the measured data from air showers above 10^{16} eV. This has been observed by multiple experiments with remarkable consistency [82–84]. Depending on the model and composition assumptions, the number of muons in simulations need to increase by $\sim$30-60% to bring the simulations into agreement with measurements. QGSJetII-04 in particular has the worst agreement with data, requiring an increase of 61% [85].

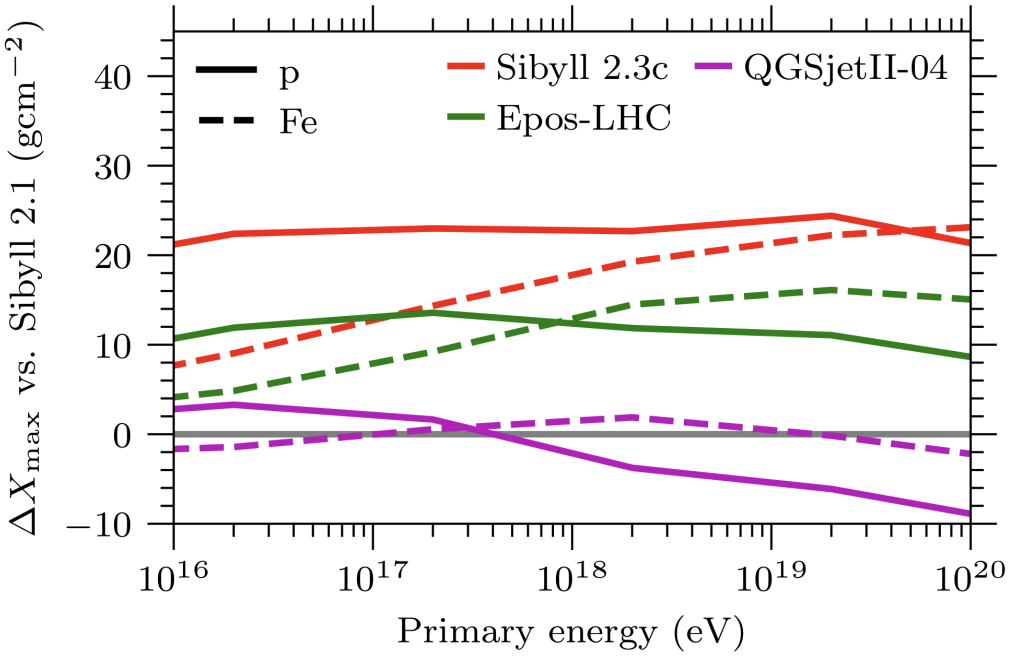


Figure 2.12. Difference in the prediction of $X_{\rm max}$ between the latest hadronic interaction models and Sibyll 2.1. The latest hadronic models generally predict a deeper shower maximum compared to Sibyll 2.1 [81].

2.5.2 Measurements

While the origin and composition of the highest energy cosmic rays are not yet understood, both Auger and TA have presented significant advances in our understanding of the mass composition of UHECRs in recent years. Shown in Figures 2.13 and 2.14 are the moments of the $X_{\rm max}$ distributions measured by Auger and TA, respectively. Auger has a higher sensitivity since it has accumulated more statistics compared to TA due to both its larger aperture and longer operation time. The differences between Auger and TA have been thoroughly investigated and it has been shown that there is no contradiction in the measurements between the two experiments [86–89].

2.5.2.1 Fluorescence Detector Measurements

One of the most reliable measurements for mass composition analysis is the elongation rate since it is independent of any hadronic interaction models. Shown in Figure 2.13 is the elongation rate measured by Auger. There is a clear break in the elongation rate at $E_0 = 10^{18.32\pm0.03}$ eV (where the composition is the lightest), above and below which it is getting heavier. The elongation rate is 77 ± 2 g/cm^2/dec, and 26 ± 2 g/cm^2/dec below and above the break, respectively [90]. The TA measure-

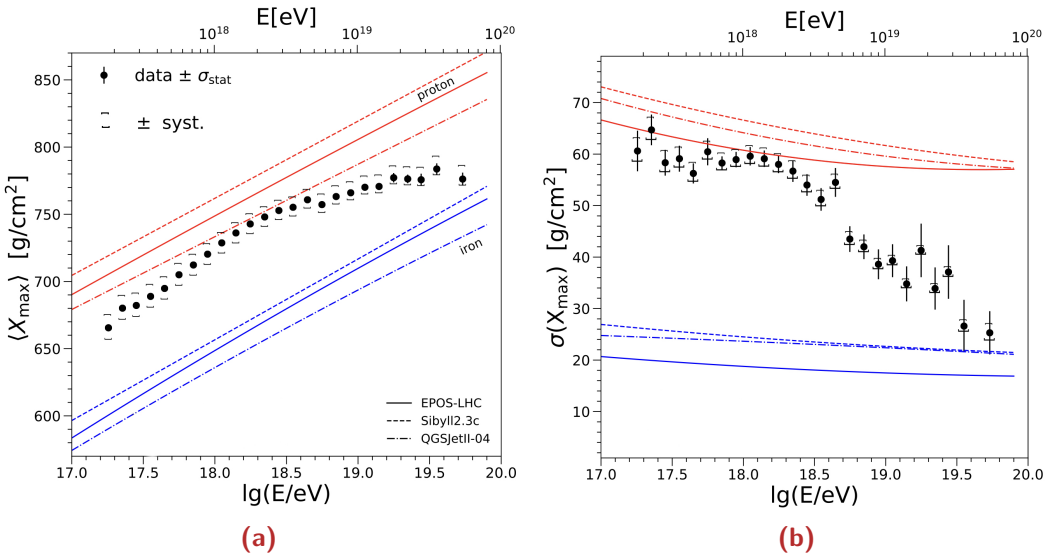


Figure 2.13. Measurements of (a) $\langle X_{\text{max}} \rangle$ and (b) $\sigma\,(X_{\text{max}})$ from Auger compared to the predictions for proton and iron nuclei of the hadronic models EPOS-LHC, Sibyll 2.3c and QGSJetII-04 [90].

ments are also compatible with a break in the elongation rate; however, a constant elongation rate cannot be excluded due to large statistical uncertainties [91].

The measurements of $\langle X_{\text{max}} \rangle$ and $\sigma\,(X_{\text{max}})$ from Auger are shown in Figure 2.13. The results suggest an increasingly heavy mass composition above E_0 from the rapid change of $\langle X_{\text{max}} \rangle$ which is in agreement with the Monte Carlo (MC) predictions of heavier nuclei. Additionally, the Auger results above $10^{18.5}$ eV show a significant decrease in $\sigma\,(X_{\text{max}})$, indicating that the mass composition mixture above the ankle becomes less dispersed and can only contain a small fraction of protons at the highest energies [90]. The measurements of $\langle X_{\text{max}} \rangle$ and $\sigma\,(X_{\text{max}})$ from TA are shown in Figure 2.14. TA follows a different analysis procedure compared with Auger, where the hadronic model predictions are folded with the detector response in order to be compared to a measured X_{max} distribution. Auger on the other hand, obtains unbiased X_{max} distributions by only selecting events with certain shower geometries. These geometries are chosen to allow a wider range of X_{max} values to be observed with full detection efficiency. This means that it is difficult to directly compare the two analyses [89]. Despite it not being possible to directly compare the X_{max} moments from the two experiments, it is possible to compare their conclusions. While Auger sees strong evidence for a heavier composition, the results from TA cannot exclude a light mass composition. Nevertheless, a mixed composition of 57% proton, 18% helium, 17% nitrogen and 8% iron (fixed composition between $10^{18.2}$ and $10^{19.1}$ eV) provides the best fit to the TA measurements [91].

In order to test the compatibility of the X_{max} moments from Auger and TA, a new technique was developed to convert the (unbiased) Auger measurements into

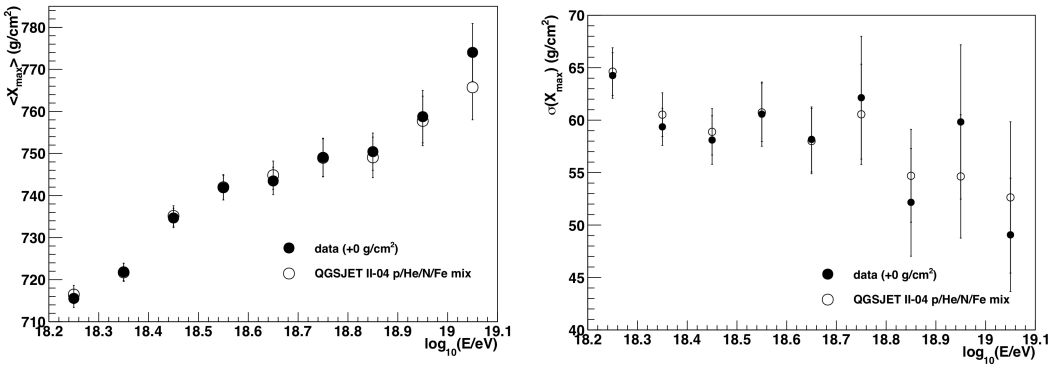


Figure 2.14. Measurements of (a) $\langle X_{\max} \rangle$ and (b) $\sigma\left(X_{\max}\right)$ from TA compared to the predictions for a mixed composition fraction of 57% proton, 18% helium, 17% nitrogen and 8% iron using QGSJetII-04 [91].

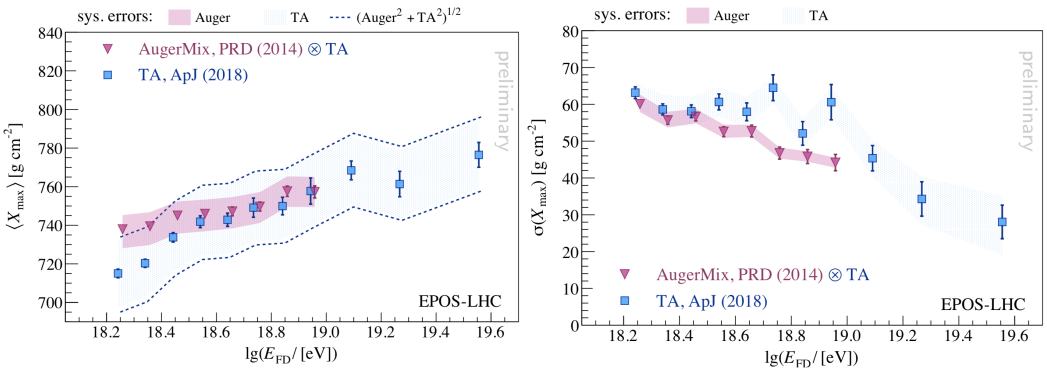


Figure 2.15. Comparison of $\langle X_{\max} \rangle$ (left) and $\sigma\left(X_{\max}\right)$ (right) between TA $X_{\max}$ moments (blue) and the moments produced by folding the Auger measurements with a MC simulation of the TA detector (pink) [89].

measurements folded with the TA detector effects. The method utilises the fitted composition fractions from the Auger measurements to produce a MC data set which is passed through a MC simulation of the TA detector. The results of the comparison are shown in Figure 2.15. The $\langle X_{\max} \rangle$ of TA and Auger agree within statistical and systematic errors, especially above $10^{18.5}$ eV. Comparison of $\sigma\left(X_{\max}\right)$ shows good agreement below $10^{18.7}$ eV. Above $10^{18.7}$ eV the statistical fluctuations of the TA measurements become too large to make meaningful comparisons with the Auger measurements. Overall there is no contradiction in the measurements between the two experiments. Future quantitative comparisons are planned to more rigorously test the compatibility between the TA and Auger $X_{\max}$ measurements [89].

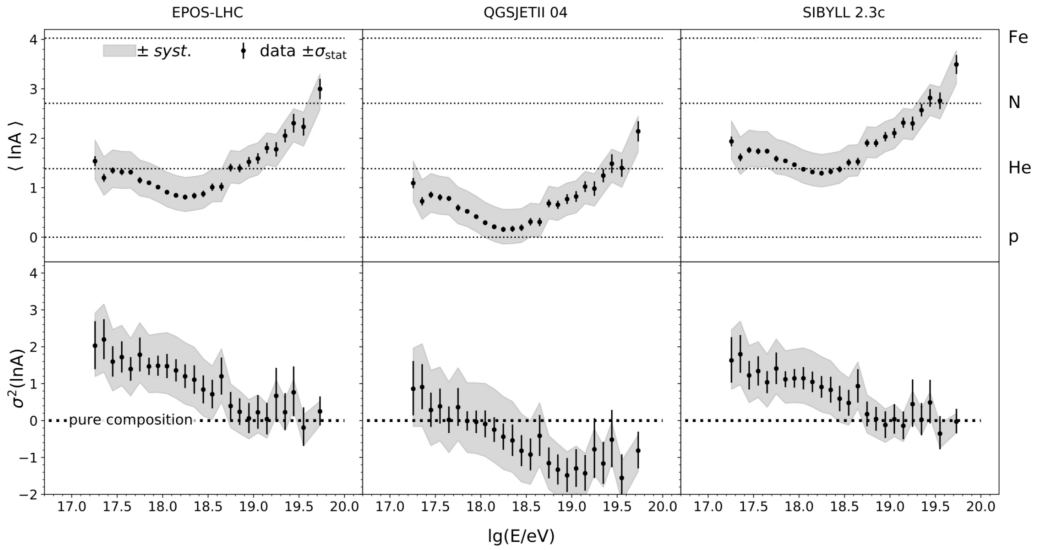


Figure 2.16. The moments of the $\ln A$ distributions converted from the $X_{\max}$ distributions measured by Auger [90].

The moments of the $X_{\max}$ distribution are related to the moments of $\ln A$ distributions (where A is atomic mass) [90, 92, 93]:

$$\langle X_{\max} \rangle = \langle X_{\max} \rangle_p + f_E \langle \ln A \rangle \tag{2.46}$$

$$\sigma^2(X_{\max}) = \left\langle \sigma^2_{\text{sh}} \right\rangle + f_E^2 \sigma^2 (\ln A) \tag{2.47}$$

where $\langle X_{\max} \rangle_p$ and $\langle \sigma^2_{\text{sh}} \rangle$ are the mean $X_{\max}$ for protons and the composition-averaged shower-to-shower fluctuations, and f_E is a parameter describing the hadronic model [93]. Using this information, the moments of the $X_{\max}$ distributions can be converted into direct measurements of mass composition. The results of this using Auger data are shown in Figure 2.16. The break at E_0 is clear in Figure 2.16 regardless of hadronic model, however the value of $\langle \ln A \rangle$ at the minimum varies from ~ 0 for QGSJetII-04 to ~ 1.4 for Sibyll 2.3c, illustrating the strong dependence on the hadronic model for the interpretation of mass composition measurements [90]. The large difference between QGSJetII-04 and the other two hadronic models; EPOS-LHC and Sibyll 2.3c in Figure 2.16 can be attributed to its significantly shallower prediction of $X_{\max}$ (see Figure 2.12). The negative variance of the mass ($\sigma^2 (\ln A) < 0$) for QGSJetII-04 is unphysical. Therefore, it can be concluded that QGSJetII-04 is not compatible with the Auger measurements.

2.5.2.2 Surface Detector Measurements

In addition to mass composition measurements with FDs, Auger has developed a method for estimating $X_{\max}$ using its SD stations. The so-called Delta method relies on information provided by the rise times of signals recorded in the SD stations [94,

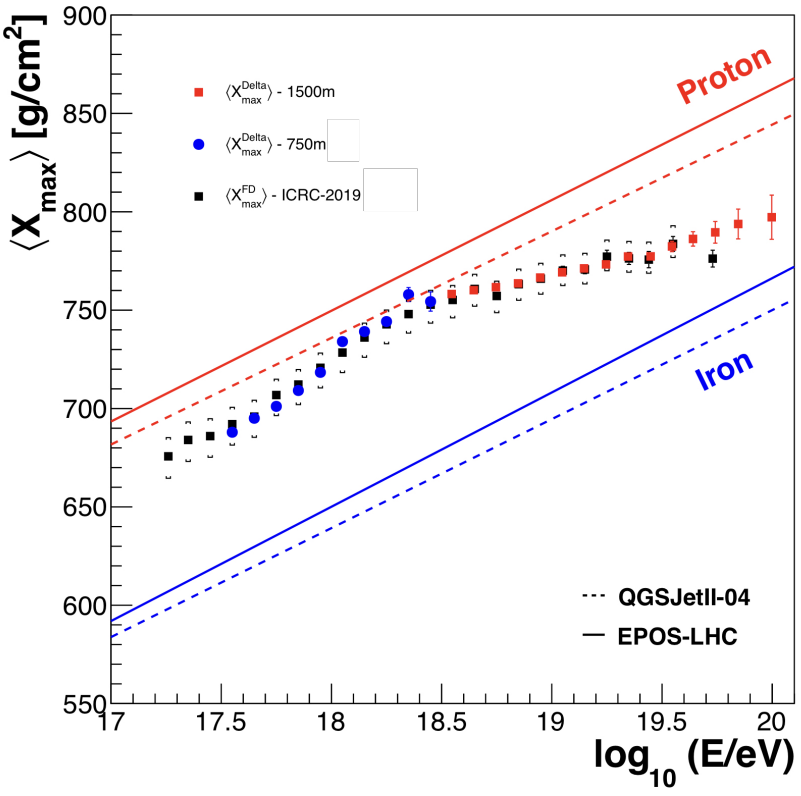


Figure 2.17. Measurements of $\langle X_{\mathrm{max}} \rangle$ from the Auger SD 1500 m and 750 m array are shown as the red and blue points, respectively. For comparison, the FD measurements from Figure 2.13a are shown as the black points. Also plotted are the predictions for proton and iron nuclei of the hadronic models EPOS-LHC and QGSJetII-04 [94].

95]. The rise time of signals in SD stations is a measure of the spread in arrival times of shower particles at the ground. Particles created at the beginning of the shower arrive first, and particles created near X_{max} arrive later. Following this idea, it can be concluded that the spread in arrival times of particles will be larger for showers with a deeper X_{max}. Consequently, the rise time of the signal measured in SD stations is correlated with X_{max}, and therefore correlated with the mass composition of the primary particle. The rise time, $t_{1/2}$ is defined as the time taken for the integrated signal to increase from 10% to 50% of its total. This definition of rise time also makes it sensitive to the muon-electron ratio of the shower since muons arrive at ground level before electrons and photons due to the higher attenuation and multiple scattering of electromagnetic particles in the atmosphere [95]. The Delta method works by characterising each event with a parameter, Δ_S which is defined as

$$\Delta_S = \frac{1}{N}\sum_{i}^{N}\frac{t_{1/2} - t_{1/2}^{\text{bench}}}{\sigma_{1/2}} \tag{2.48}$$

where the sum is over the N stations which recorded signals in the event, $t_{1/2}^{\text{bench}}$ is a benchmark rise time which represents the average rise time as a function of core distance and zenith angle, and $\sigma_{1/2}$ is the rise time uncertainty [95]. Using high quality hybrid events measured by both the FD and SD, Δ_S is calibrated to X_{max} as measured by the FD [94, 95]. The results of this analysis are directly compared with the Auger FD measurements in Figure 2.17. The results show good agreement, validating the method [94].

2.5.3 Implications

Measurements of mass composition at the highest energies are limited by statistics as discussed in Section 2.2. While Auger has found evidence for a composition progressively heavier than pure proton above the ankle, other experiments such as TA cannot exclude a pure proton composition. The HiRes (Section 2.6.5) experiment also found evidence for a protonic composition [96]. As discussed in Section 2.2 there is a clear suppression in the energy spectrum at the highest energies. Given that the TA results cannot exclude a light composition at the highest energies, the suppression can still be reasonably interpreted as the GZK effect. However, the Auger measurements indicate a heavier component in the mass composition at the highest energies. Given Auger's interpretation of the mass results, the energy of cosmic rays may be limited by a maximum rigidity or by photo-disintegration, however, it is still unclear which energy-loss mechanism dominates. Above 10^{19} eV, these two models produce significantly different predictions of mass composition. In the maximum rigidity scenario, a heavier composition is expected since the gyro-radii of protons is too large. In the photo-disintegration scenario, on the other hand, a large fraction

of protons is expected. In order to distinguish these two scenarios, a significant increase in composition-sensitive data is required to allow a reliable interpretation of the observed X_{max} distributions, as well as further cross-checks and extension of hadronic models [97, 98].

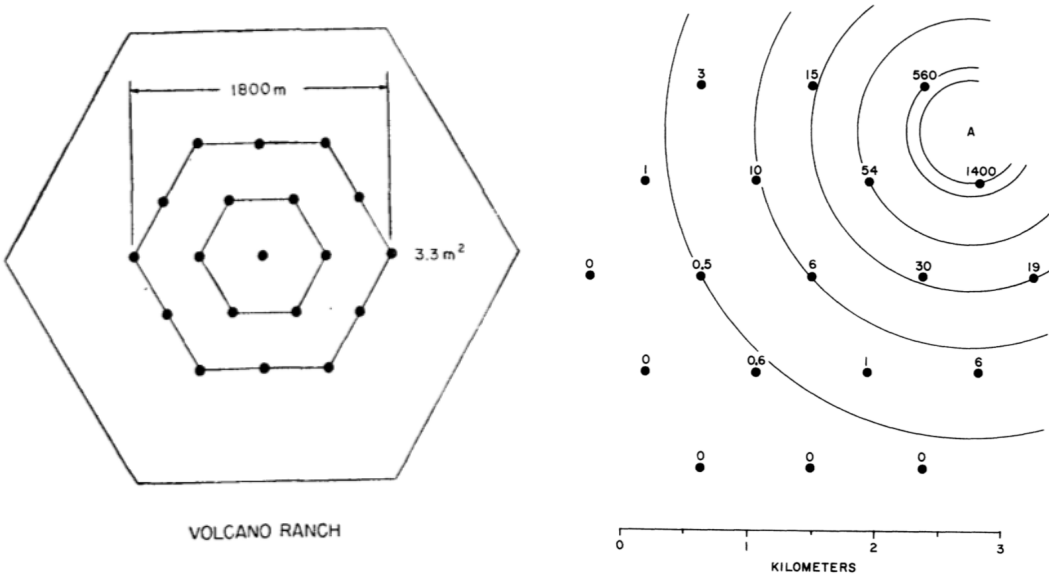
2.6 Cosmic Ray Experiments

The work in this thesis is based on two cosmic ray experiments; namely, the Pierre Auger Observatory (Auger) and the Fluorescence detector Array of Single-pixel Telescopes (FAST). These experiments are discussed in detail in Chapters 3 and 6, respectively. Presented in this section are brief descriptions of other notable cosmic ray experiments.

2.6.1 Volcano Ranch

Volcano Ranch, built in Albuquerque, New Mexico, was the first giant air shower array. It was operated between 1959 and 1963 by a research group led by John Linsley and Bruno Rossi from the Massachusetts Institute of Technology [14, 99, 100]. The Volcano Ranch array consisted of 19 triangular arranged 3.3 m^2 scintillation counters, initially with a total area of $\sim$2 km^2 during its first run in 1959-1960. The array spacing was then doubled for its second run from 1960 to produce an area of $\sim$8 km^2, as shown in Figure 2.18a [101].

The cosmic ray energy spectrum beyond 10^{18} eV was first measured by the Volcano Ranch array. These measurements included evidence for a flattening in the spectrum corresponding to the ankle 20 years before it was confirmed conclusively [103]. Data from the Volcano Ranch array was also used in the first studies of the arrival directions of the highest energy cosmic rays, which found no evidence for anisotropy in arrival directions [103]. In February 1962, the highest energy event ever recorded at the time was measured by the array with an energy of 1.4×10^{20} eV. The density map of the event is shown in Figure 2.18b. To this day it remains one of the highest energy events ever detected [14, 102]. In 1976, detection of atmospheric fluorescence and Cherenkov light emission from air showers was made using an optical system by a group from the University of Utah. The optical system was operated in coincidence with the Volcano Ranch array in order to determine if it was possible to measure air showers using such optical systems, and if the measured signals agreed with the expectation from the coincident measurements made by the surface array [104]. The success of these measurements led to the development of the Fly's Eye detector (Section 2.6.5).



(a) (b)

Figure 2.18. (a) The layout of the Volcano Ranch array, showing the position of the 19 scintillation detectors (black dots) in their original configuration as well as the size of the expanded array, which corresponds to the outer hexagon [101] and (b) the density map of the event above 10^{20} eV. The numbers correspond to the shower density at each detector (particles/m^2) and point 'A' corresponds to the estimated shower core [102].

2.6.2 Haverah Park

The Haverah Park extensive air shower array was built as a collaborative experiment between Universities of Durham, Leeds, Nottingham and London (Imperial College). The array operated between 1968 and 1987 and consisted of water Cherenkov detectors distributed over an area of $\sim$12 km^2 with an irregular spacing due to restricted land access [105, 106]. The experiment originally consisted of only a 500 m array of four detectors which recorded its first air showers in December 1962 [107]. These detectors are shown in Figure 2.19 as A1-A4. Each of the four detectors of area 34 m^2 consisted of 15 individual Cherenkov detectors made out of galvanised steel tanks with an area of 2.29 m^2 and a height of 1.2 m. Each tank was instrumented by a single 5-inch PMT [105–107]. In addition to the 500 m array, six sub-arrays comprising of four 13.5 m^2 detectors surrounded the 500 m array at $\sim$2 km from its centre, shown as B to G in Figure 2.19. From May 1980, several plastic scintillation detectors were operated at Haverah Park in order to perform cross-calibration with Volcano Ranch and Yakutsk. They found excellent agreement in measurements of the lateral distribution function (LDF) and energy calibration between the three experiments [108]. The success of the Haverah Park experiment and its use of water Cherenkov detectors paved the way for current experiments like Auger, which later used a similar design for its SD.

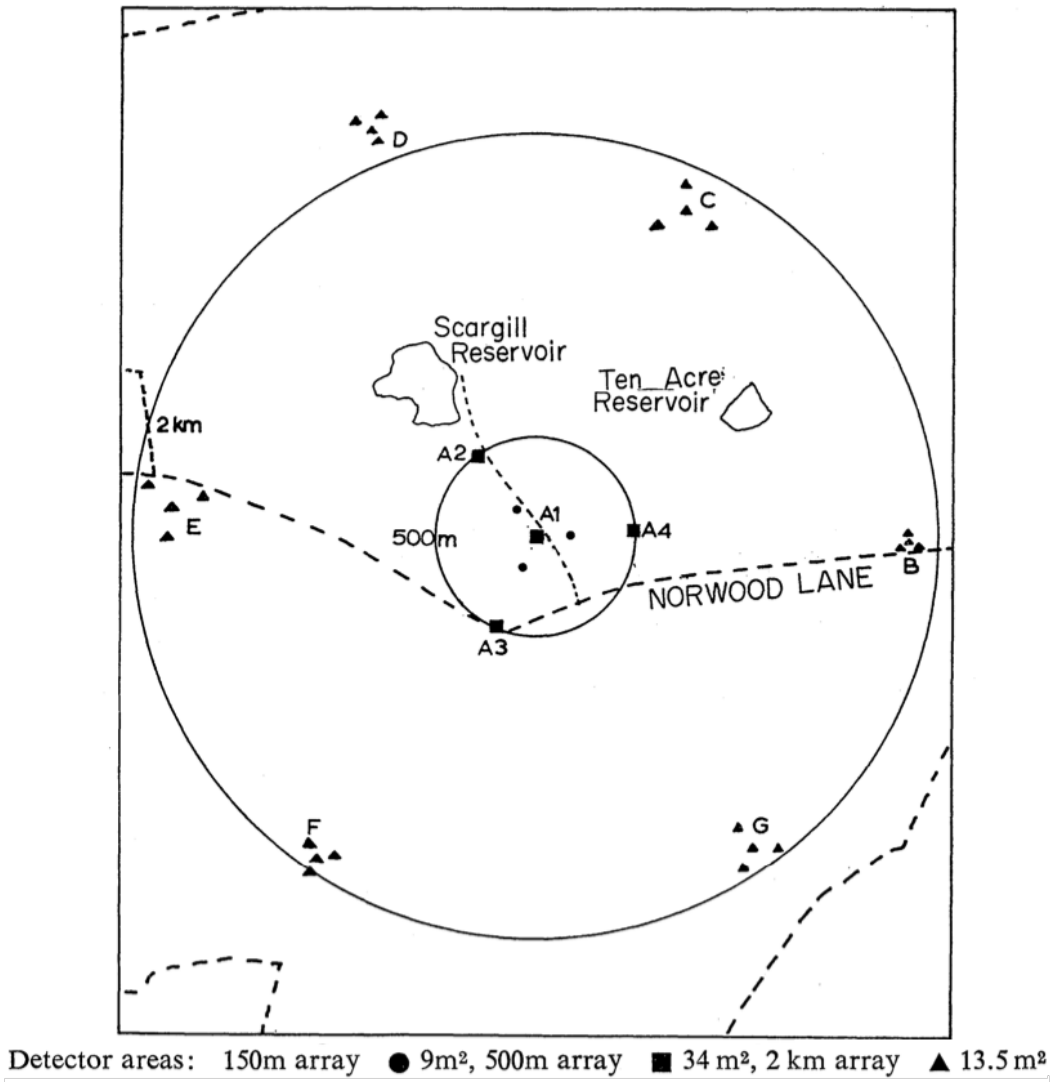


Figure 2.19. This map shows the layout of the Haverah Park array [105].

2.6.3 Sydney University Giant Air-shower Recorder (SUGAR)

The Sydney University Giant Air-shower Recorder (SUGAR) was the first giant air shower array built in the Southern Hemisphere, offering a unique vantage point from which to study the arrival direction of cosmic rays in contrast to the array's Northern Hemisphere counterparts [109]. SUGAR was built by the University of Sydney near the town of Narrabri, New South Wales, Australia and it operated from 1968 to 1979. The array consisted of 54 pairs of 6 m^2 liquid scintillation detectors distributed over an area of more than 60 km^2. The large size of the array provided the ability to detect very high energy showers at the expense of resolution. Each of the detectors was buried 1.7 m below ground level to reduce the background rate and the detector pairs were separated by 50 m in the North-South direction [110, 111]. The array was therefore sensitive to the muon component of air showers rather than the electromagnetic component. The large spacing of the detectors (typically $\sim$1.6 km) eventually proved too great for reliable measurements of air showers due to the small number of triggered stations per event, however the novel approach of designing fully autonomous stations with their own local power source proved a huge success and the concept would later be used by modern arrays like the Telescope Array and the Pierre Auger Observatory [14].

2.6.4 Yakutsk

The Yakutsk array is located near Yakutsk, Russia and has been used to study cosmic rays since 1970 [112, 113]. The array originally consisted of scintillators covering an area of 17 km^2 from 1974 to 1990. The array was later contracted (now covering 8 km^2) in order to take detailed measurements of the lateral distribution of air showers. The layout of the array is shown in Figure 2.20. Currently the array consists of 58 scintillator stations, 48 Cherenkov light detectors, four underground muon detectors and six 32 MHz radio detectors [114]. The Cherenkov light detectors measure air-Cherenkov light produced by showers using upwards-facing PMTs. The detectors are separated into two sub-arrays, C_1 and C_2 with 500 m and 50-200 m spacing, respectively. The scintillators can be directly calibrated using the calorimetric energy measurements of the Cherenkov detectors [113]. Yakutsk has made contributions to the measurement of the cosmic ray energy spectrum in the range of 10^{15}-10^{20} eV as well as cosmic ray composition and proton-air cross section measurements [112]. The Yakutsk array is currently undergoing a modernisation process in order to improve measurement accuracy [114].

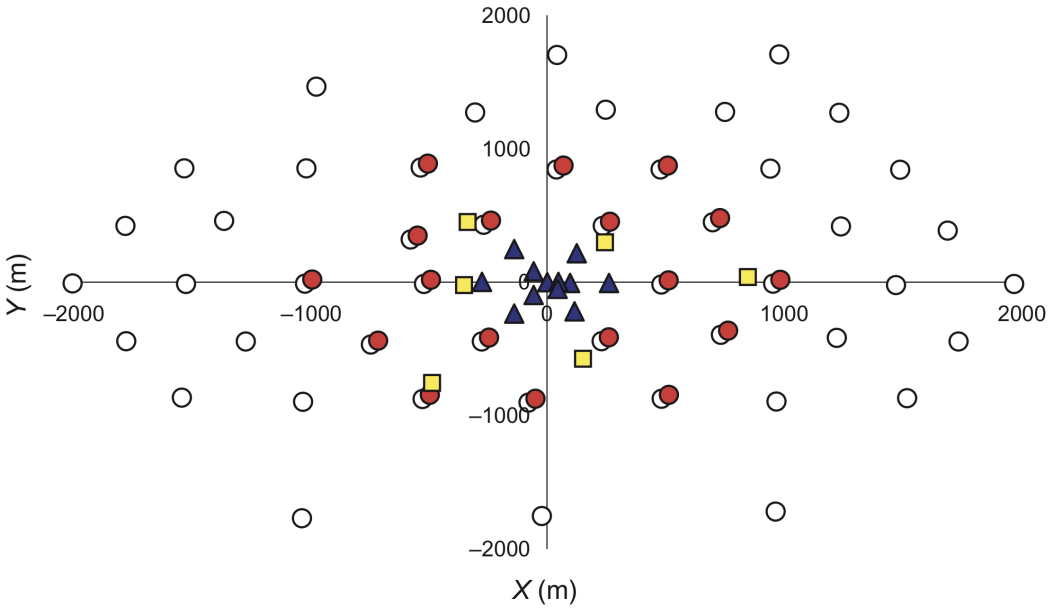


Figure 2.20. Layout of the Yakutsk array. Scintillators are shown as open black circles, Cherenkov light detectors of the C_1 and C_2 subset are shown as red circles and blue triangles, respectively. Muon detectors are shown as yellow squares. [113].

2.6.5 The Fly's Eye and HiRes

The Fly's Eye detector was located at the U.S. Army's Dugway Proving Ground in Western Utah, USA. The first detector (Fly's Eye I) began operation in 1981, followed by a second detector (Fly's Eye II) in 1986. The Fly's Eye was the first cosmic ray detector to successfully use the air fluorescence technique to measure air showers [115]. Fly's Eye I comprised of 67 1.5 m diameter mirrors mounted in cylindrical steel drums, providing full-sky coverage as shown in Figure 2.21. The sky was imaged by an array of either 12 or 14 PMTs in each telescope, and a total of 880 PMTs. Fly's Eye II was located 3.3 km from Fly's Eye I and consisted of 36 mirrors and a total of 464 PMTs, viewing half of the sky in the direction of Fly's Eye I. This allowed for the stereoscopic analysis of air showers, providing a significant improvement in reconstruction accuracy [115, 116].

The Fly's Eye made significant contributions to the understanding of the cosmic ray energy spectrum, including measurements of the ankle at $\sim 3 \times 10^{18}$ eV as well as evidence for the GZK cut-off. The highest energy cosmic ray to date was measured by Fly's Eye I on the 15$^{\text{th}}$ of October 1991 at an energy of $(3.2 \pm 0.38$ (statistical) $\pm$ 0.85 (systematic)$) \times 10^{20}$ eV. The Fly's Eye also contributed to the understanding of cosmic ray mass composition using measurements of the depth of shower maximum, X_{max} [116, 118]. The Fly's Eye experiment ended its operation in July 1992 [116].

Figure 2.21. Image of the Fly's Eye I detector with mirrors mounted in steel drums and PMT arrays [117].

The High Resolution Fly's Eye (HiRes) experiment, which consisted of two FD stations (HiRes-I and HiRes-II) 12.6 km apart, was operated from 1998 to 2006 at the existing Fly's Eye site. The goal of HiRes was to increase the aperture by an order of magnitude above 10 EeV and to improve the reconstruction resolution [118, 119]. The HiRes telescopes achieved a factor of 7 increase in the signal-to-noise ratio by increasing the number of PMTs observing the same field of view (FoV) with a 16 $\times$ 16 array of PMTs, and the mirror diameter from 1.5 to 2 m. HiRes-I consisted of 22 mirrors with an elevation FoV from 3° to 17° and HiRes-II consisted of 43 mirrors, providing a larger elevation FoV from 3° to 30° [119, 120]. HiRes was the first to observe the GZK cut-off in the cosmic ray energy spectrum, at an energy of 6×10^{19} eV [121, 122]. The observed energy spectrum shown in Figure 2.22 clearly shows the ankle and the GZK cut-off. HiRes also made significant progress in mass composition measurements using X_{max} above 10^{18} eV which are consistent with a proton-dominated composition [96, 122].

2.6.6 Akeno Giant Air Shower Array (AGASA)

The Akeno Giant Air-Shower Array (AGASA) was located near the village of Akeno in Japan, and covered an area of 100 km^2, making it the largest cosmic ray observatory during its operation from 1990 to 2004 [123–125]. The Akeno Giant Air-Shower Array (AGASA) consisted of 111 2.2 m^2 scintillation detectors with a spacing of 1 km. Additionally, 27 of the scintillation detector sites also contained a muon

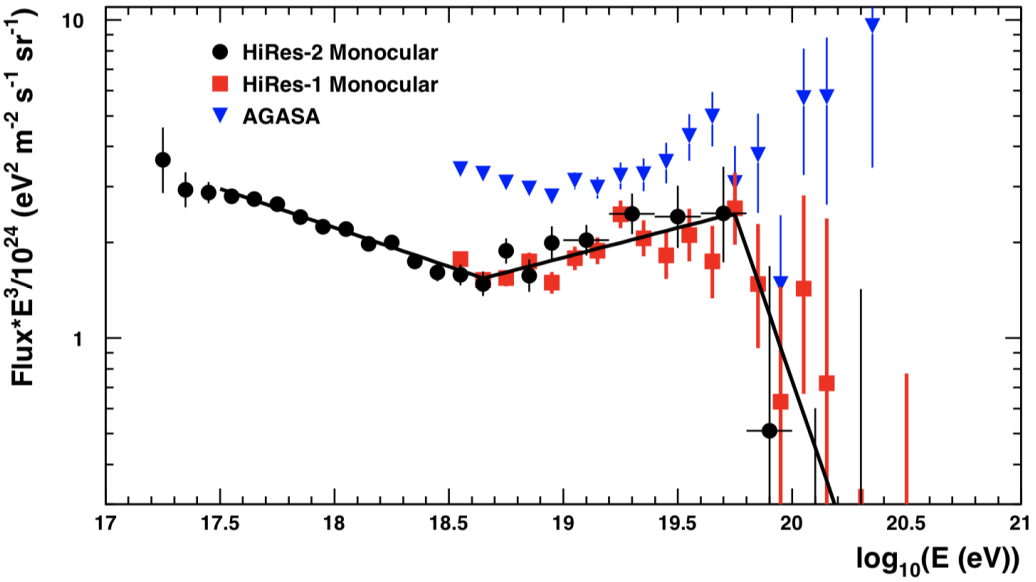


Figure 2.22. Cosmic ray energy spectrum measured by HiRes compared to AGASA. There is clear evidence for the ankle and the GZK cut-off in the HiRes measurements [121].

detector [123]. The array originally consisted of four branches with separate trigger systems, but was merged in 1995, providing a significant increase in its effective area. In total, AGASA reached an exposure of 5.8×10^{16} m^2 s sr with almost 1000 events detected above 10^{19} eV [125]. The most notable result obtained by AGASA was the observation of 11 events above the GZK energy limit, providing an energy spectrum inconsistent with the theoretical GZK cut-off, and with the energy spectrum of other experiments [124–126]. The disagreement between HiRes and AGASA inspired a new generation of detectors, the Pierre Auger Observatory (Auger) and the Telescope Array (TA).

2.6.7 The Telescope Array Project (TA)

The Telescope Array (TA) experiment is currently the second largest operational cosmic ray experiment, and the largest in the Northern Hemisphere. It is analogous to Auger, as it also employs a hybrid detection technique comprising an SD array and surrounding FD stations. TA is situated near the town of Delta in Utah, USA and has been operating since 2007 [127, 128].

The TA SD array comprises 507 plastic scintillators, each with a collecting area of 3 m^2. These detectors are arranged on a 1.2 km square grid and cover a total area of $\sim$700 km^2, as shown in Figures 2.23 and 2.24 [129]. The SDs measure the arrival time, and the lateral distribution and density of extensive air shower particles at the ground. In addition to the SD stations, a total of 38 FD telescopes overlook

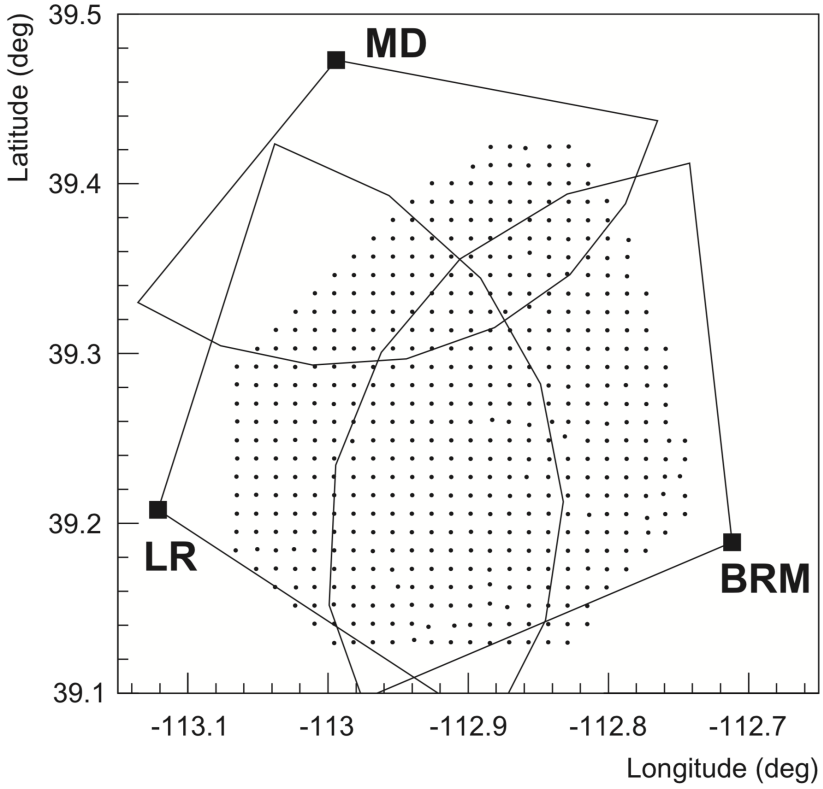


Figure 2.23. A map of the TA site. The black points show the locations of the current TA SD stations with a spacing of 1.2 km. Also shown are the locations of the FD sites. The black lines show the approximate FoV of the FD telescopes for 10^{19} eV showers. Note that the MD site utilises re-purposed HiRes telescopes, and BRM and LR house the new TA telescopes [130].

the array. These telescopes are installed at three separate sites on the edge of the array, Black Rock Mesa (BRM), Long Ridge (LR), and Middle Drum (MD) which are shown in Figure 2.23. The MD site houses 14 fluorescence telescopes which have been re-purposed from the HiRes detectors, while the BRM and LR sites each house 12 new fluorescence telescopes. Each FD telescope consists of a spherical primary mirror with a diameter of 3.3 m and a PMT camera comprising 256 PMTs arranged in a 16×16 hexagonal grid. The FoV of the camera is 15° in elevation and 18° in azimuth. Both of the new FD stations provide a total FoV from 3° to 33° in elevation and 108° in azimuth [128].

The Telescope Array Low-energy Extension (TALE) provides an additional 103 SD stations with various spacings, as well as an additional 10 FD telescopes located at the TA MD station. The additional FD telescopes view higher elevations, from 31° to 59°. TALE lowers the TA energy threshold down to $10^{16.5}$ eV with reasonable resolution and an expected event rate of 5000 hybrid events per year. While at high energies the hybrid combination of fluorescence light and scintillation detectors is the preferred method, at lower energies TALE uses air-Cherenkov light instead of fluorescence light for air-shower observations [131, 132].

The flux suppression expected from the GZK limit has been observed by TA at an energy of 5.4×10^{19} eV [9, 133]. Another result of the TA experiment is evidence for a "hotspot" of cosmic rays above 57 EeV, positioned about 19° off of the super-galactic plane. The probability of such a hotspot appearing by chance in an isotropic cosmic ray sky was initially reported as 3.7×10^{-4} (3.4σ) [64]. The most recent results using the TA 11-year sample find a local significance of 5.1σ and the chance probability in an isotropic sky to be 2.1×10^{-3} (2.9σ) [65].

TA is currently undergoing a significant upgrade called TAx4, which includes the extension of its SD array to almost 3000 km^2, comparable to the size of Auger [134]. An additional 500 SD stations are being deployed to the north-east and south-east of the existing array with a 2 km spacing. Two additional FD stations will also be built at MD (4 telescopes) and BRM (8 telescopes). These upgrades are shown in Figure 2.24. The upgrade aims to increase statistics at the highest energies by a factor of 4 [135].

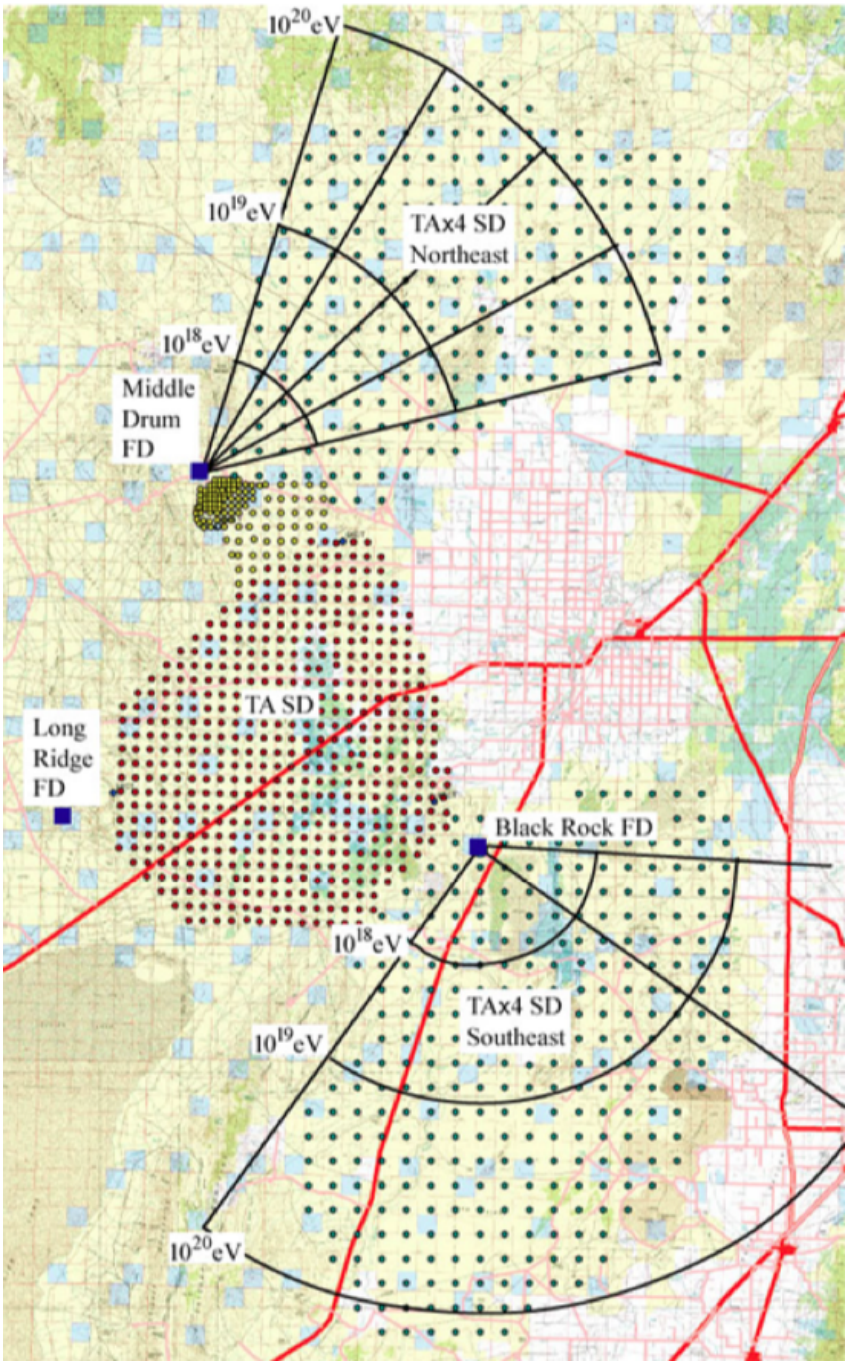


Figure 2.24. An overview of the TA site including the ongoing upgrades. The red circles (middle) show the location of the current TA SD stations with a spacing of 1.2 km. The yellow circles to the north of the main array show the location of TALE SD stations [131, 132]. The green circles in the north and south correspond to the planned locations of the TAx4 SD stations with a spacing of 2.08 km. The black lines show the planned FoV of TAx4 FD telescopes [135].

3 The Pierre Auger Observatory

The Pierre Auger Observatory (Auger) is located near the town of Malargüe in Mendoza, Argentina. It is currently the largest cosmic ray detector in the world. The observatory was conceived of in 1991 by physicists James Cronin and Alan Watson for the purpose of understanding the origin and nature of the highest energy cosmic rays. Following the results of previous experiments, it was clear that only a very large detector would reach the exposure needed to answer these questions. With a current total of over 14 years of data acquisition, Auger has reached a massive exposure of approximately 80,000 km^2 sr yr using a large Surface Detector (SD) and a set of Fluorescence Detector (FD) telescopes [22]. The SD consists of 1660 water-Cherenkov detectors with a spacing of 1500 m, distributed over an area of ~ 3000 km^2. The mean altitude of the array is $\sim$1400 m (equivalent to a vertical atmospheric depth of 875 g/cm^2). The SD array is overlooked by four FD sites, each consisting of six fluorescence telescopes [136]. The SD and FD provide two independent detection techniques which are used in conjunction to observe air showers. This enables cross-checks and higher statistical confidence in measurements. The combination of the SD and FD is called a hybrid observatory and provides a significant advantage in the reconstruction resolution of air shower events measured by Auger [136]. The layout of the SD array with the surrounding FD sites is shown in Figure 3.1.

3.1 Surface Detector

The SD is an array of 1660 water-Cherenkov detectors placed in a triangular arrangement with a spacing of 1500 m [137]. In addition to the regular array, an infill array of 61 stations with a spacing of 750 m is located near the Coihueco FD site as a part of the Auger Muons and Infill for the Ground Array (AMIGA) extension (a low energy extension of the observatory) [136]. The layout of the SD stations (including the 750 m array stations) is shown in Figure 3.1.

Each SD station consists of a water tank made from polyethylene with a height of 1.55 m and a diameter of 3.6 m. The tank is filled to a height of 1.2 m with purified water. Mounted inside the tank are three 9-inch diameter photomultiplier tubes (PMTs) which record the Cherenkov light emitted by relativistic charged particles as

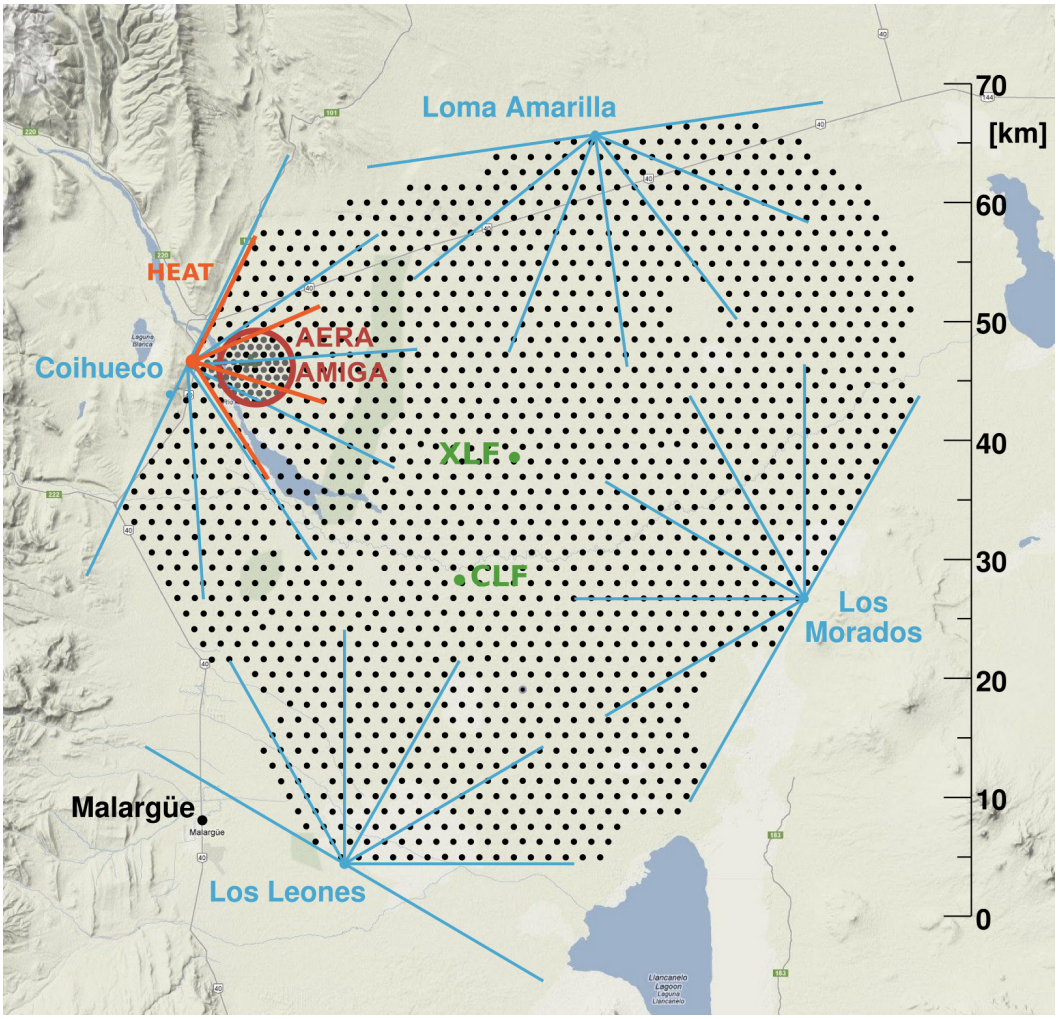


Figure 3.1. The layout of the Pierre Auger Observatory (Auger). Each black dot indicates a single SD station. The blue lines indicate the fields of view (FoV) of the FD telescopes which overlook the SD array [136].

Figure 3.2. A surface detector station near-by the Los Leones FD site. The solar panel and communications antenna are visible. The tank was designed to blend in with the surrounding landscape. Image was taken by the Author.

they traverse the water. An inner lining provides diffuse reflectivity of the Cherenkov light [137]. A photograph of a single SD station near the Los Leones FD site is shown in Figure 3.2. High energy photons are converted into electron-position pairs through pair production in the water, making the SD sensitive to high energy photons as well as charged particles and muons. The Cherenkov light is measured in a unit known as a vertical equivalent muon (VEM). A VEM is defined as the signal produced by a muon passing through the middle of the tank on a vertical trajectory [136, 137]. The SD stations are self-contained with power supplied by a 10 W battery which is charged by solar panels attached to the top of the tank. On board electronics process the signals from air shower events. The signals are digitised and transmitted to the Central Data Acquisition System (CDAS) via the on-board communications antennae [137]. A schematic diagram of a SD station is shown in Figure 3.3.

3.1.1 Station Calibration

The measurement of the average signal produced by vertical through-going atmospheric background muons, called the VEM, is used to calibrate the SD stations [139]. The aim of the calibration is to determine the integrated charge (electronic signal

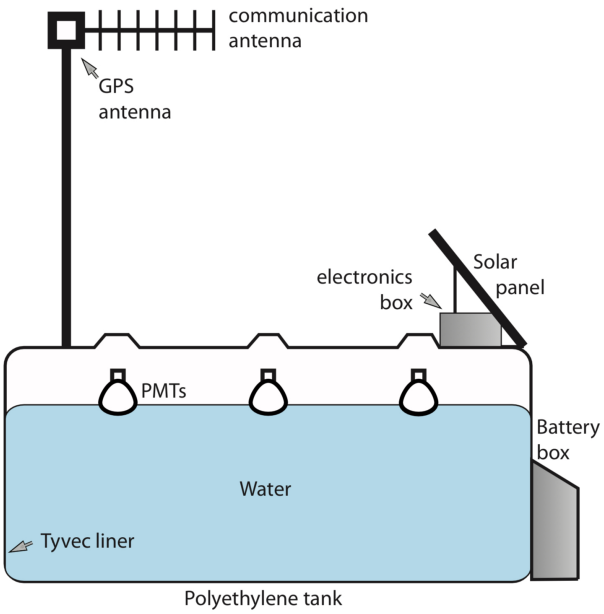


Figure 3.3. A schematic diagram of a water-Cherenkov detector, showing its main components [138].

units) produced by a 1 VEM signal. Each station is constantly monitoring the integrated charge and the amplitude (pulse height) of background signals. It is important to monitor in real-time in order to maintain a consistent trigger, and hence constant energy threshold across the array. The integrated charges produced in the PMT electronics by atmospheric background muons are collected, and a distribution of the signals is produced. A large number of particles can be obtained due to the extremely high background rate of muons. The signal due to the vertical and through-going background muons produces a distinct peak in the charge distribution. This peak is used to convert the electronic analogue-to-digital converter (ADC) signal into VEM [136, 139]. Shown in Figure 3.4 are example charge and pulse height distributions from an SD station. The first peak is caused by low-energy particles, and the second peak by omnidirectional muons. The red peak corresponds to the vertical and through-going muons of interest. The known ratio of the peaks in the charge distribution (Figure 3.4(left)) of the omnidirectional and vertical muons can be used to convert measured signals to VEM. An additional, more accurate procedure is performed offline during reconstruction using the signals measured over a short period of time immediately prior to an event. This involves a robust method for finding the peak corresponding to the VEM by fitting a second-order polynomial with negative curvature [139].

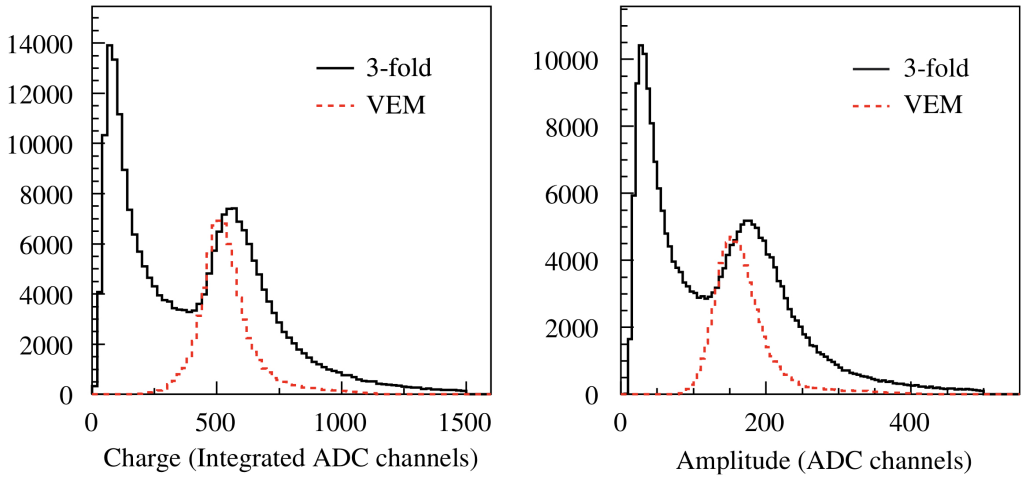


Figure 3.4. Charge (left) and pulse height (right) histograms produced by an SD station. The black distributions represent the sum from a 3-fold coincidence trigger of the three PMTs in the tank. The first peak is caused by low energy particles, and the second peak by omnidirectional muons. The dashed red distribution represents the signal produced by vertical through-going muons as determined by an external muon detector trigger [139].

3.1.2 Surface Detector Triggers

The SD trigger algorithms are divided into a hierarchy of five levels. The first two levels of triggers, T1 and T2 are formed locally at individual SD stations in real time. The third level, T3 is formed by combinations of T2 station triggers at the Central Data Acquisition System (CDAS), while the fourth and fifth levels, T4 and T5 are offline physics-based triggers used to select real events from the background. The SD triggers described in this section are the basis for the analysis presented in Chapter 4.

3.1.2.1 Station Level Triggers

At the level of a single SD station, the T1 and T2 triggers are used to identify patterns in the signal traces recorded by the three PMTs in the tank [140]. There are four types of local station triggers:

1. **Threshold (Th)**
   A simple threshold trigger which requires 1 bin of the trace to be above 1.75 I_{VEM} and 3.2 I_{VEM} for a T1 and T2 level trigger, respectively. Here, I_{VEM} is the peak amplitude of a pulse from a vertical muon. The background trigger rate of the T1 level trigger is 100 Hz. The T2 trigger reduces this to 20 Hz in order to cope with the limited bandwidth between individual SD stations and CDAS.

2. **Time-over-threshold (ToT)**
   The ToT trigger requires 13 signal bins within a 3 μs (120 bins) sliding window to be above a threshold of 0.2 I_{VEM}. ToT triggers are automatically promoted to the T2 level due to the low rate of background particles mimicking the required signal.

3. **Time-over-threshold de-convoluted (ToTd)**
   The ToTd trigger is the first of the new triggers implemented in 2013. It is a more complicated algorithm which applies a deconvolution to remove the exponential tail of Cherenkov light pulses produced by electromagnetic particles in the tank. Pulses from muons or other shower particles will have a sharp rise caused by direct Cherenkov light, followed by a longer tail caused by light being reflected by the inner liner of the tank. The deconvolved signal in trace bin i is given by

$$D_i = \frac{S_i - S_{i-1}e^{-\Delta t/\tau}}{1 - e^{-\Delta t/\tau}} \tag{3.1}$$

   where S_i is the measured signal, Δt is the width of a single time bin (25 ns) and τ is the average decay time of the Cherenkov light (67 ns). The standard ToT algorithm is then applied to the deconvolved trace and the trigger is automatically promoted to the T2 level if it passes the required threshold of 0.2 I_{VEM}. These requirements result in a trigger rate of 0.3 Hz.

4. **Multiplicity of positive steps (MoPS)**
   The MoPS trigger is the second of the new triggers implemented in 2013. The algorithm is completely independent of the other trigger types. It counts the number of consecutive positive-going steps in the signal, as well as keeping track of the amplitude, j of the increases at each step. For each group of consecutive increasing signal bins, j is required to fall between 3 and 31 in order to be above the average noise, but below that of the average consecutive bin increase due to a vertical muon. Each time such a group passes this condition, it is counted towards the multiplicity, m. The trigger is accepted when $m > 4$ within a 120-bin window. Like the ToT and ToTd algorithms, MoPS triggers are automatically promoted to the T2 level. These requirements result in a trigger rate of 0.3 Hz.

These triggers and their requirements are summarised in Figure 3.5. Signals which only pass the threshold trigger generally resemble short duration spikes. This makes the threshold trigger particularly effective at detecting inclined air showers where the showers have travelled through a greater atmospheric depth. This has the effect of attenuating the electromagnetic component of the shower such that the signal is dominated by muons. The ToT trigger, on the other hand, is better at detecting

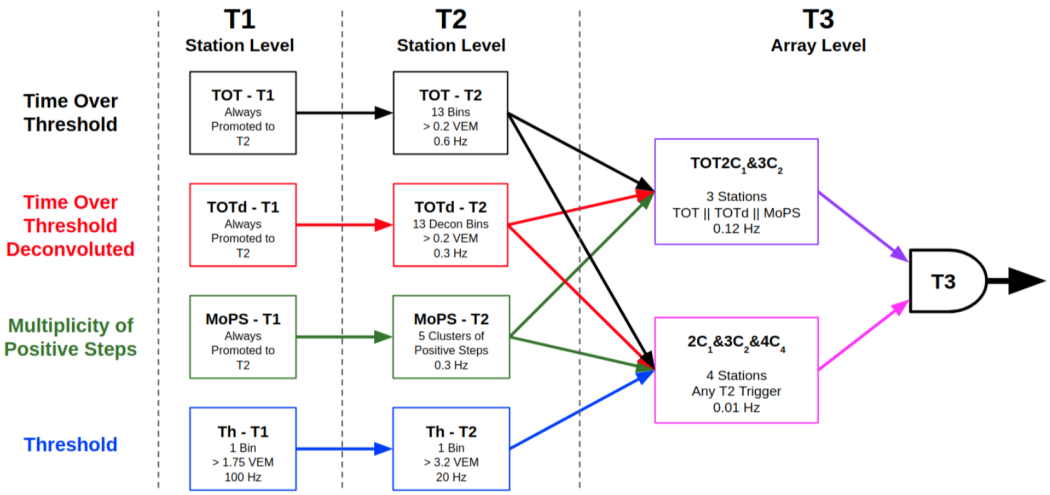


Figure 3.5. A diagram summarising the types of station level and array level triggers as well as their requirements [141].

pulses with smaller amplitudes, but longer durations, complimenting the threshold trigger. The new triggers, ToTd and MoPS significantly increase the sensitivity of the SD to low energy showers. This is because the typical background muon signal appears as a signal spike with a width of only one or two bins. While such a signal will cause random Th triggers, these signals do not influence the ToTd or MoPS triggers since additional shower particles are required to form a trigger. This enables the reduction of the energy threshold of the array, while maintaining an acceptable background trigger rate [136]. The station level triggers are sent to CDAS in real time where the data are collated.

3.1.2.2 Array Level Trigger

The third trigger level, T3 is formed based on spatial and temporal combinations of T2 triggered stations. The trigger is comprised of two modes which both have requirements on the compactness of T2 stations. The compactness is determined by the number of crowns over which the T2 stations are spread. The first crown, C_1 consists of a station's closest six neighbours. The second crown, C_2 refers to the next hexagon of stations concentrically outwards from the first crown, and so on as shown in Figure 3.6. The two modes are:

1. **ToT2C$_1$&3C$_2$**
   The first of the T3 trigger modes is the 3-fold mode (TOT2C$_1$&3C$_2$), which requires at least three ToT triggered stations in a strict compact configuration with at least two stations contained within one crown (2C$_1$) and all three within two crowns (3C$_2$) as illustrated in Figure 3.6. In this case a ToT trigger can also include either of the new triggers ToTd or MoPS.

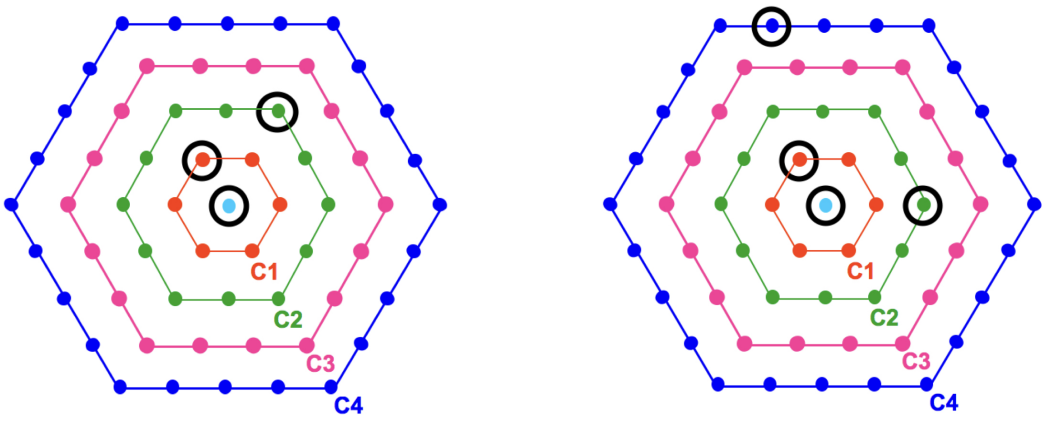


Figure 3.6. The first four crowns of SD stations are shown with a possible configuration for the $2C_1\&3C_2$ trigger (left) and the $2C_1\&3C_2\&4C_4$ trigger (right) [140].

2. **$2C_1\&3C_2\&4C_4$**
   The second mode is the 4-fold mode ($2C_1\&3C_2\&4C_4$), which allows any type of T2 trigger, but requires at least four stations where the first three stations have the same requirement as the 3-fold mode, and all four must be within four crowns.

When a T3 trigger of the surface array is formed, all signals passing the T2 requirements are sent to CDAS, as well as any T1 triggers in time-coincidence (30 μs) of the T3 trigger.

3.1.2.3 Event Selection

The fourth and fifth triggers, T4 and T5 are performed offline as part of the reconstruction procedure. The T4 trigger is the first of the event selection, or physics triggers. It is used to identify real events and to reject noise events from the T3 data set. A T4 trigger can be formed in two ways:

1. **3ToT**
   The 3ToT trigger requires at least 3 T2-ToT triggers in an explicit triangular pattern, where at least two of the triangle's sides are defined by adjacent stations (spacing of 1500 m or 750 m). The possible configurations are shown in Figure 3.7a. The trigger also requires that the trigger times of the stations are consistent with a shower front moving at the speed of light; that is the difference in trigger times of the stations must be less than their distance apart divided by the speed of light [140]. The 3ToT trigger is dominant for vertical showers with zenith angles $< 60°$.

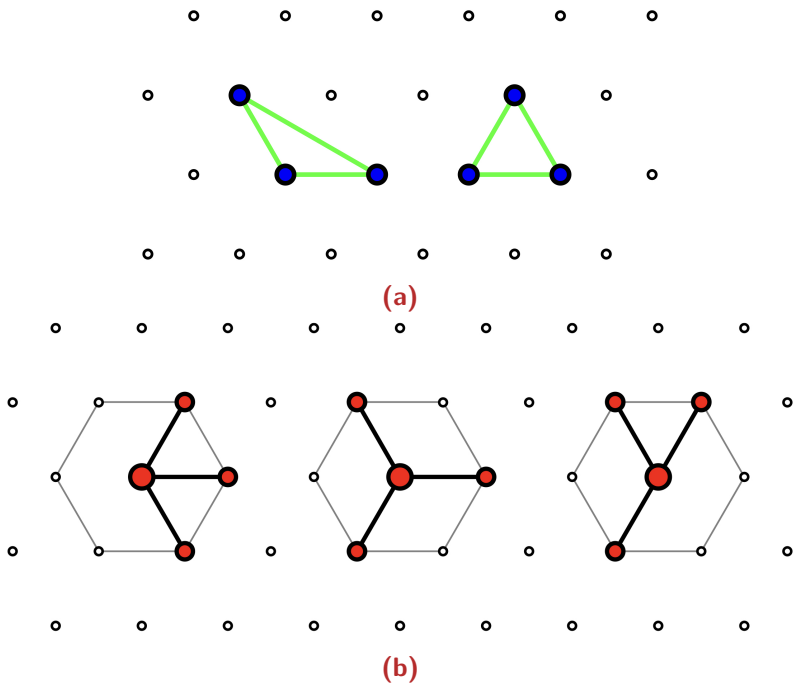


Figure 3.7. (a) Possible 3ToT station configurations and (b) possible 4C1 station configurations.

2. **4C1**
   In order to form a 4C1 trigger, at least four stations are required, but without any condition of the type of T2 trigger. This means the 4C1 trigger is most effective at zenith angles $> 60°$, where the threshold triggers dominate due to atmospheric attenuation of the electromagnetic component of the shower. The stations must all be contained within one crown, as shown in Figure 3.7b. Like the 3ToT trigger, the 4C1 trigger also requires that the trigger times of the stations are consistent with a shower front moving at the speed of light [140].

The T5 trigger is applied to reject events which are not fully contained inside the array. This includes showers which fall near the edge of the array such that part of the shower is not detected. The trigger requires the hottest station (the station with the largest signal) to be fully surrounded by six working stations. This also has the effect of rejecting events which contain stations that are not functioning.

3.2 Fluorescence Detector

Auger has a total of 27 fluorescence telescopes which are located at four sites on the perimeter of the SD array, namely Los Leones, Los Morados, Loma Amarilla and Coihueco. A photograph of the Los Leones FD building is shown in Figure 3.8. With the exception of Coihueco, each site contains six telescopes. Each telescope has

Figure 3.8. A photograph of the FD site building at Los Leones [136].

a FoV of $30^\circ \times 30^\circ$ in azimuth and elevation, combining to give a total azimuthal FoV of 180° at each site, which is represented by the blue lines in Figure 3.1 [136]. Coihueco also contains an additional three telescopes called the High-elevation Auger telescopes (HEAT), which observe higher elevations. HEAT is discussed further in Section 3.2.3.

The FD observes the longitudinal development of air showers as they pass through the atmosphere. These observations utilise measurements of the fluorescence light emitted isotropically along the path of the shower. However, only a small fraction of the shower energy is emitted as fluorescence light. The fluorescence yield, defined as the number of photons produced by a shower particle per unit energy deposit has been measured in the laboratory by experiments such as AIRFLY to very high precision [142, 143]. Based on these precise measurements, the fluorescence light emitted by an extensive air shower (EAS) can be used to estimate the energy of the shower. The atmosphere therefore functions as an energy calorimeter, making FD measurements of EAS energy very reliable [136, 144, 145]. This estimate of the energy is known as the calorimetric energy. A correction for "invisible energy" must be considered to obtain the total energy of the shower. This is discussed further in Section 3.5.2.7.

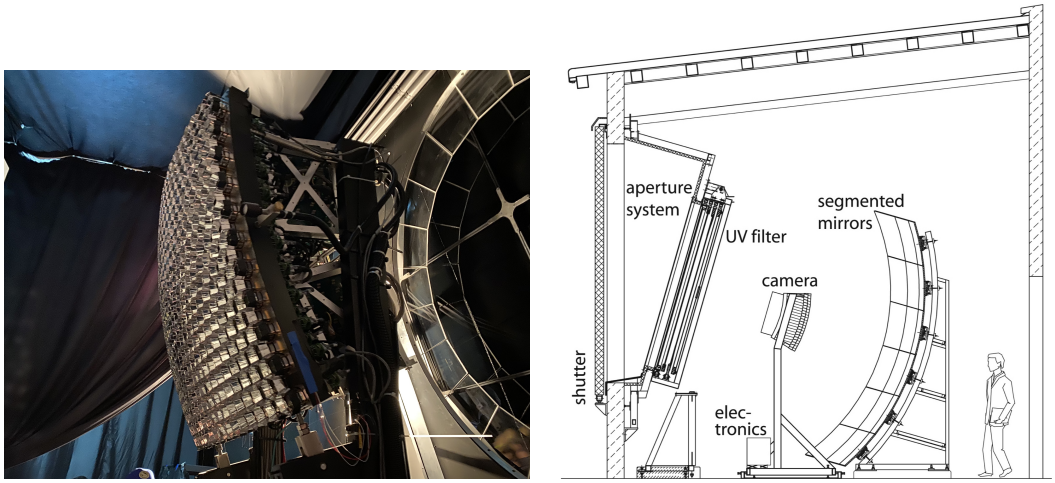


Figure 3.9. (a) A photograph of a telescope camera at Coihueco taken by the Author and (b) A schematic diagram illustrating the primary components of an FD telescope [144].

3.2.1 Hardware and Electronics

Each of the FD telescopes are housed in a climate-controlled building which contains all of the hardware, electronics and communications equipment required for data acquisition [136, 144]. A schematic diagram of the FD camera is shown in Figure 3.9b. An ultra-violet (UV) filter is mounted to the telescope's aperture to remove the background light from stars and artificial light sources which are not associated with the fluorescence light emitted by air showers. This significantly increases the signal-to-noise ratio of the measurements. The light passing through the filter is focused onto the camera using a 13 m^2 segmented spherical mirror. An annular corrector ring is mounted on the inside of the filter to correct for spherical aberrations, thus keeping the spot size on the camera within the design requirement of 0.5°. The camera consists of 440 hexagonal PMTs arranged on a 20 $\times$ 22 grid, with each pixel viewing a 1.5° diameter region of the sky. A photograph of a FD camera at Coihueco is shown in Figure 3.9a. Each PMT pixel is surrounded by light collectors known as *Mercedes stars* which circumvent the dead space between pixels where light would otherwise be lost. The light collection efficiency is improved from 70% to 94% with the use of the Mercedes stars [144]. An electronics distribution board supplies each group of 44 PMTs with a low voltage (LV) and high voltage (HV) source as well as acquiring the signal from the PMTs [136]. The front-end electronics filter and digitize the PMT signals at 10 MHz (100 ns time bins). The signals are filtered with a 10 MHz, fourth-order Bessel anti-aliasing filter to remove high frequency noise. Dedicated electronics boards also provide two hardware triggers for individual pixel triggers and shower track identification.

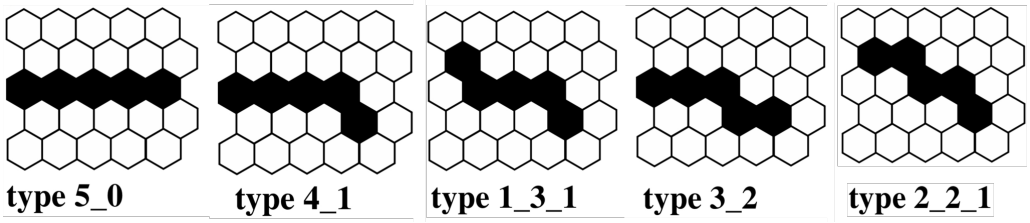


Figure 3.10. The fundamental triggered pixel patterns which are considered in track identification by the SLT. Adapted from [146].

3.2.1.1 Telescope Trigger System

A series of hardware and software triggers process the signals recorded by the PMTs of the FD camera. The four trigger types for the FD are:

1. **First Level Trigger (FLT)**
   The FLT identifies individual pixel triggers using an integrated signal threshold cut. The integrated signal is determined by a moving "boxcar" window. The width of the window is fixed to 1 μs (10 time bins) and the integrated signal threshold is dynamically adjusted in order to maintain a constant trigger rate of 100 Hz [144]. This method significantly improves the signal-to-noise ratio [144, 146].

2. **Second Level Trigger (SLT)**
   The triggered pixels are combined by the SLT based on a pattern recognition search which attempts to identify one of the fundamental trigger patterns shown in Figure 3.10. Any rotation or mirroring of the patterns is allowed. These fundamental patterns correspond to the possible segments of a straight track of triggered pixels across the FD camera. The algorithm requires at least four of the five pixels from the pattern to trigger. The trigger rate of the SLT is between 0.1 and 10 Hz per telescope [144].

3. **Third Level Trigger (TLT)**
   The TLT is a two-stage software-based trigger algorithm which is designed to reject noise events without the costly readout of the entire set of ADC traces. Noise triggers can be caused by lightning, muons entering the glass of a PMT, or by random pixel triggers [136, 144]. The first stage of the TLT is designed to reject the lightning-based events by determining the number of simultaneous FLT pixels within the same time bin. Cuts are placed on the multiplicity of such triggers based on Monte Carlo (MC) simulations [147]. The second stage of the TLT is based on a rudimentary reconstruction of the shower-detector plane (SDP) (see Section 3.5.2) which searches for tracks of spatial and time-coincident pixels and rejects pixels with large deviations.

Based on MC simulations, 95% of background events are correctly rejected by the TLT and less than 1% of real events are rejected [147]. The trigger rate is reduced to 0.01 Hz by the TLT [144].

4. **Hybrid Trigger (T3)**
   When a TLT is formed it is sent to CDAS, which subsequently sends a signal to the SD stations to return any type of trigger which occurred within 20 μs of the FD trigger time [136]. Triggers from all telescopes at an individual FD building are combined and a first estimate of the reconstruction is performed online within 1 second before the trigger is sent to CDAS [136]. The trigger rate is 0.02 Hz per FD building. The T3 trigger is extremely important since the SD will not trigger independently at the array level for low energy showers which have only triggered one or two stations. However, as discussed in Section 3.5.3, even a single SD station is sufficient for hybrid reconstruction.

3.2.2 Telescope Calibration

The calibration of the FD telescopes is a critical step towards the reconstruction of an EAS [148–150]. In order to convert the electronic signal of ADC counts to a flux of photons at the aperture of the telescope, it must first be calibrated to determine the response of each PMT to a given flux of photons. The calibration is performed in three ways; a semi-regular absolute calibration based on a fixed wavelength light source placed at the telescope aperture, a relative multi-wavelength calibration, and nightly relative calibrations which track changes in the detector response over time [149]. In addition to these measurements, an independent measurement using a portable vertical laser can be used as a cross check of the absolute calibration [144, 151].

3.2.2.1 Absolute Calibration

The absolute calibration is performed using a 1.4 m deep, 2.5 m diameter "drum" mounted at the telescope's aperture, as shown in Figure 3.11. The drum consists of a pulsed UV LED light source at a fixed wavelength of 375 nm. The light source is mounted inside the drum along with reflective surfaces which scatter the light through a thin sheet of Teflon on the front face of the drum. This provides a diffuse light source at the telescope aperture which uniformly illuminates all 440 pixels of the camera. The drum is calibrated in the laboratory to a precision of 4% or better, in order to ensure it provides a known, uniform light source. The average response of the FD is 5 photons per ADC count. The full calibration procedure takes into account the filter transmittance, mirror reflectivity, light collection efficiency, quantum efficiency of the PMTs, and the electronics [136, 144].

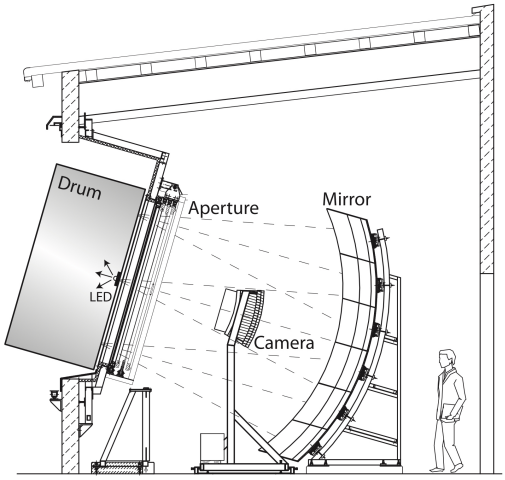


Figure 3.11. A schematic diagram of the FD telescope with the drum attached to the aperture for the absolute calibration procedure [144].

3.2.2.2 Multi-wavelength Calibration

Initial measurements of the wavelength dependence of the telescope calibration involved the use of the drum and a Xenon flasher mounted to the back of the drum as a replacement for the fixed wavelength light source. The Xenon flasher was capable of providing five different wavelengths of light at 320, 337, 355, 380 and 405 nm [149, 150]. Updated measurements using a monochromator in addition to a Xenon flasher provided additional measurements in steps of 5 nm from 280 nm to 440 nm in wavelength [148]. The relative spectral calibration efficiencies for the average of telescopes 4 and 5 at Los Morados are shown in Figure 3.12 and compared with the five wavelength measurements. The shape of the spectral response is primarily due to the spectral dependence of the PMT quantum efficiency and the UV filter transmission [144]. The overall uncertainty in the spectral response of the FD telescopes is 1.5%. The relative spectral calibration changes the average reconstructed energy of showers by less than 3% and has almost no effect on the depth of shower maximum [148].

3.2.2.3 Relative Calibration

In addition to the absolute calibrations and the relative spectral calibrations, nightly relative calibrations provide an ongoing measurement of changes in the telescope response over time due to hardware changes and seasonal effects. This is also important since dust and dirt collect on the mirror and filter surface in-between cleaning, which changes their response to a given flux of photons. The relative calibrations are performed during an observing night using a series of three UV flashers (Cal A, Cal B and Cal C) as shown in Figure 3.13. The flashers are mounted

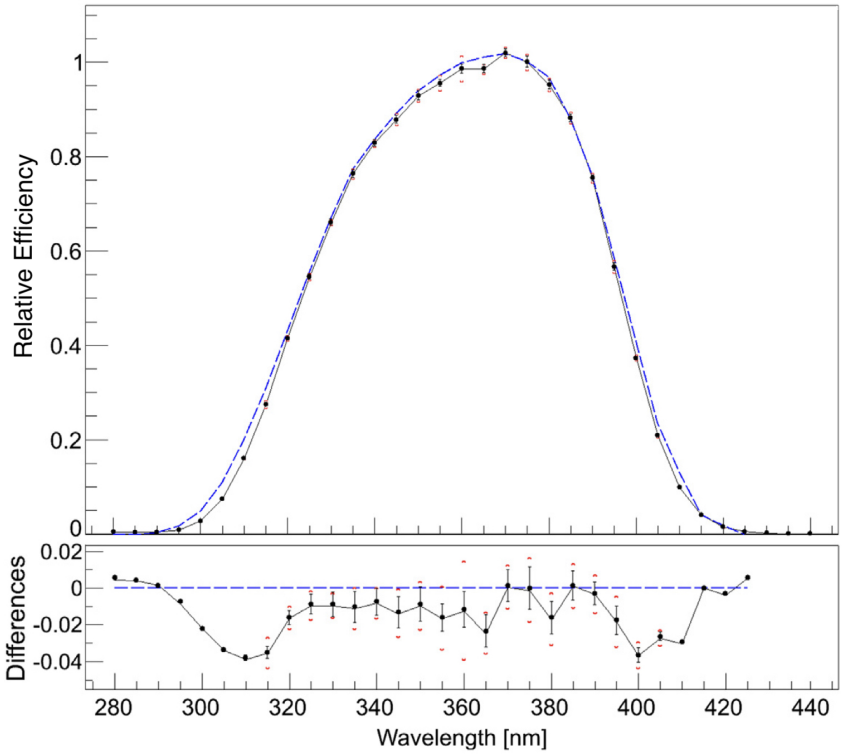


Figure 3.12. The average efficiency of Los Morados telescopes 4 and 5 as a function of wavelength relative to the efficiency at 375 nm (black points). Also shown in blue is the original spectral dependent measurements [148].

on the centre of the mirror, on the side of the camera and outside the telescope aperture. The first position, Cal A, contains a light source of 470 nm located at the centre of the mirror. It directly illuminates the camera in order to independently measure the PMT response [144, 149]. Cal A operates before and after each observing night, as well as periodically during the night. Cal B consists of two Xenon flash lamps mounted on the sides of the camera body. They illuminate the mirror and enable the combined response of the mirror and the camera to be measured. Cal C is mounted on the outside of the telescope aperture, and the light is directed towards a reflective surface that diffuses the light through the filter. This enables an end to end relative calibration of the entire optical system [144].

3.2.3 High-elevation Auger telescopes (HEAT)

In addition to the six standard telescopes at each FD site, the High-elevation Auger telescopes (HEAT) provides a low energy extension of the FD to complement the 750 m array of the SD. HEAT consists of three FD telescopes with variable tilt and is located at the Coihueco FD site, overlooking the nearby 750 m array. HEAT can be configured to extend the Coihueco FD elevation FoV from 30° to 60° in "upwards" mode, or to provide an overlapping FoV in "downwards" mode for cross-calibration of the telescopes. The HEAT telescopes utilise the standard FD optical

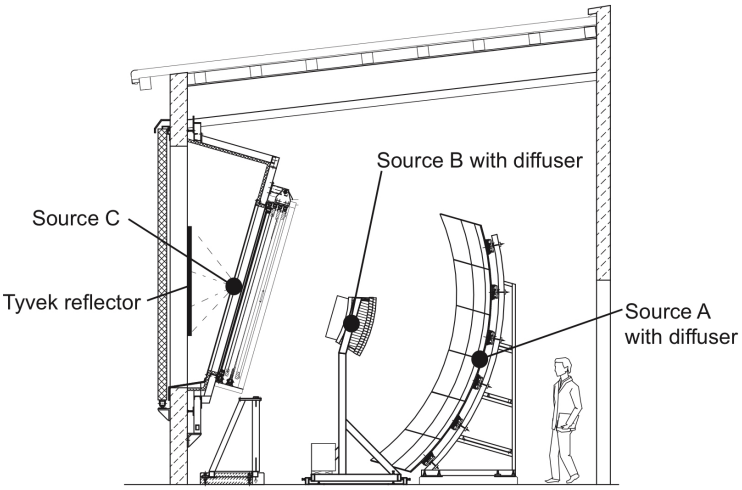


Figure 3.13. A schematic diagram of the positions of the light sources used for the relative calibration procedure [144].

design, however, the electronics were redesigned to be more sensitive to lower energy showers detected closer to the telescope, with larger angular velocities. This requires a faster sampling rate of 20 MHz compared to the 10 MHz of the standard FD electronics. The FoV of HEAT and Coihueco is combined into an additional virtual FD called HEAT-Coihueco (HeCo). It is able to observe high quality events which are used in studies of the energy spectrum and mass composition [136, 152].

3.3 Atmospheric Monitoring

As the medium through which air showers form and develop, the atmosphere serves as an essential component of a ground based cosmic ray detector such as Auger. The FD measurements in particular, depend on the assumptions made about the atmospheric conditions. It is important to quantify the fraction of light scattered and attenuated by the atmosphere on its path to the telescopes in order to accurately reconstruct the shower energy and depth of maximum, X_{max}. Auger uses several atmospheric monitoring instruments to measure the state of the atmosphere and its effect on air shower measurements.

3.3.1 Central Laser Facility and eXtreme Laser Facility

Auger operates two vertical laser facilities which are located near the centre of the SD array, where they are visible by at least one telescope from each FD site. The Central Laser Facility (CLF) has been in operation at Auger since the observatory was commissioned in 2004, while the eXtreme Laser Facility (XLF) was built later in 2008 in order to service the Loma Amarilla FD site, which is the furthest from the CLF.

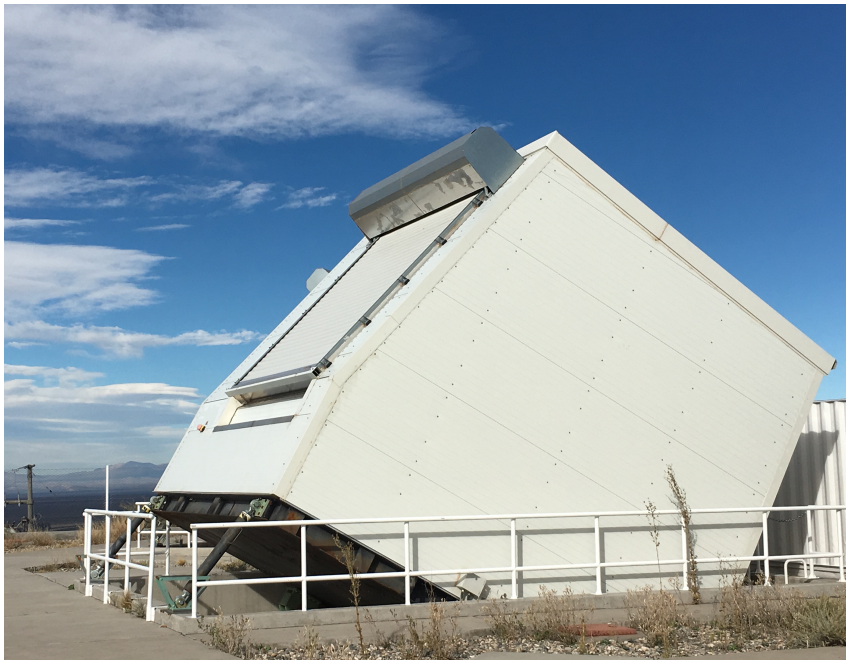


Figure 3.14. A photograph of a HEAT telescope in upwards mode.

The CLF and XLF both house a 6.5 mJ UV (355 nm wavelength) laser which fires 50 vertical pulses every 15 minutes [136, 153, 154]. The laser pulses are measured from each of the FD sites and an average hourly light profile can be derived from the photons arriving at the detector as a function of the height above the laser. The aerosol content of the atmosphere can be inferred from the comparison of the hourly light profiles to that of a "reference" night using the so-called data normalised method [155]. A reference night is chosen each year as the night with the lowest atmospheric aerosol content, and is assumed to be aerosol free [154, 156]. Hourly aerosol measurements are vital for the accurate and precise reconstruction of the shower energy and X_{max} using the FD [155, 157]. Although designed primarily to measure atmospheric aerosol content, the two laser facilities can also detect cloud directly overhead the laser and along the path between the laser and the FD telescope [154].

3.3.2 Lidars

Located at each FD site is an elastic backscatter lidar which is used to monitor cloud cover above the observatory. Each lidar consists of a laser of wavelength 351 nm, and three 80 cm mirrors, each with a PMT for light collection [136, 158–160]. Each lidar automatically scans up to 45° from the vertical. The return signal from the laser is processed to determine the distance and density of scattering regions such as clouds. An example lidar scan in Figure 3.16 shows the signature of a cloud with a base height of ~ 4 km above ground level [154].

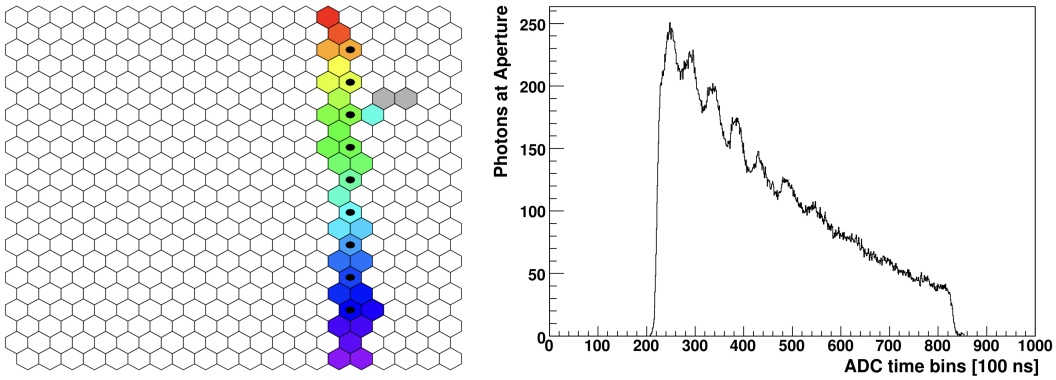


Figure 3.15. (a) An example of the laser track across the Los Leones FD site, 26 km away, and (b) an example of the average light profile seen from Los Leones. Bumps in the light profile are due to the pixelation of the camera [155].

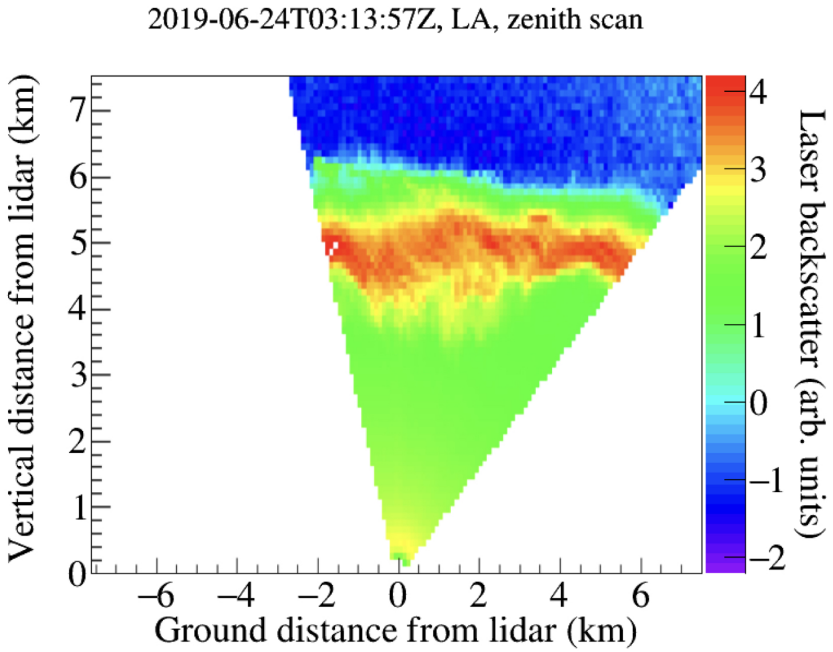


Figure 3.16. Example of an FD lidar scan [154].

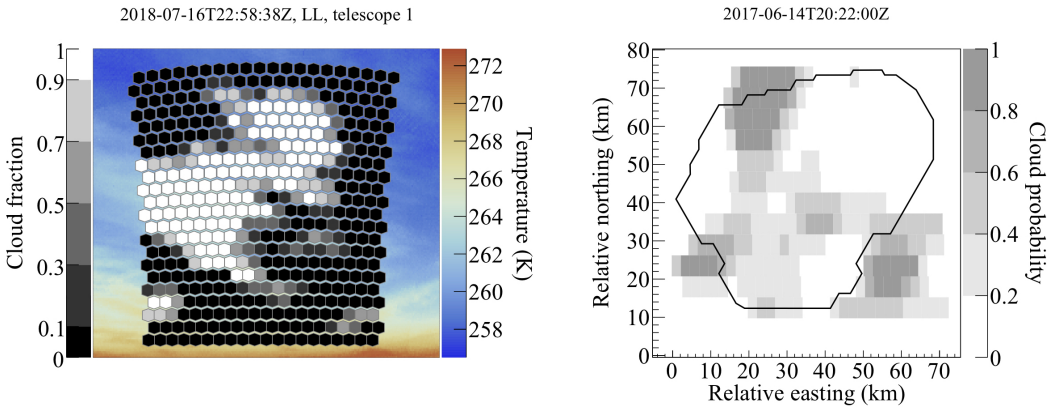


Figure 3.17. Examples of a cloud camera image with cloud fraction masked to the pixels of the FD camera (left), and a satellite cloud probability map from GOES data (right) [154].

3.3.3 Cloud Cameras

Infrared cloud cameras are mounted on the roof of each FD building, providing an image of the FD FoV every 5 minutes during data acquisition. The cameras are sensitive to wavelengths between 8 μm and 14 μm. The cloud camera images are analysed to produce a cloud fraction for each pixel of the FD camera based on the higher temperature of clouds compared to the background sky in the sensitive wavelength band. An example image is shown in Figure 3.17(left). In addition to the FoV scans, the cloud cameras also produce full sky scans every 15 minutes [154].

3.3.4 External Monitoring Tools

While Auger operates its own atmospheric monitoring instruments, the observatory also relies on several external monitoring instruments and tools such as satellites and atmospheric models. The Geostationary Operational Environmental Satellite system (GOES) is used to produce cloud probability maps over the array every 30 minutes, and the results are considered during reconstruction as part of the determination of whether an event is affected by cloud [161]. An example cloud probability map is shown in Figure 3.17(right). Radiosondes were intermittently launched at the observatory site up until the end of 2008 to study properties of the atmosphere, however in 2011 these measurements were replaced by the Global Data Assimilation System (GDAS), which is a global atmospheric model based on various meteorological measurements [162]. GDAS provides descriptions of the molecular atmosphere every three hours on a 1° grid in latitude and longitude. The nearest GDAS grid point is located near the north-eastern edge of the array. A detailed study performed to test the differences between GDAS and on-site weather balloon measurements between 2005 and 2008 has shown good agreement [162].

3.4 The Offline Software Framework

In order to analyse data collected by the observatory, a general-purpose software framework called Offline was developed. It is implemented using the C++ programming language to take advantage of its object-oriented features and for compatibility with common open source analysis tools. The Offline framework provides a sophisticated infrastructure to support the broad range of computational tasks which are required for data analysis procedures. A standard set of reconstruction and simulation modules are provided, along with various physics tools and utilities [136]. The framework can be separated into four distinct areas:

3.4.1 Central Configuration

The Offline framework uses a Central Configuration system which points the framework components and modules to the location of their configuration data [136]. This is achieved using an XML file, which can be passed to the Offline application at run time [163, 164].

3.4.2 Processing Modules and Run Control

Each task can be separated into a sequence of steps which each complete a part of the analysis procedure. The processing modules are self-contained algorithms which allow for easy collaboration and modification. This structure allows for modules to easily be switched to provide alternate algorithms or implementations. The processing modules are easily assembled and given sequencing instructions using an XML steering file [163, 164]. The basic structure of such a file is shown in Code 3.1.

3.4.3 Event Structure

The event data structure contains all data for reading and writing any information which may change between events. This includes all raw, calibrated, reconstructed and MC data. Processing modules access this data through the hierarchy shown in Figure 3.18a. Following the processing of an event, the information from its analysis is written to disk using the `ROOT` scientific analysis framework [165].

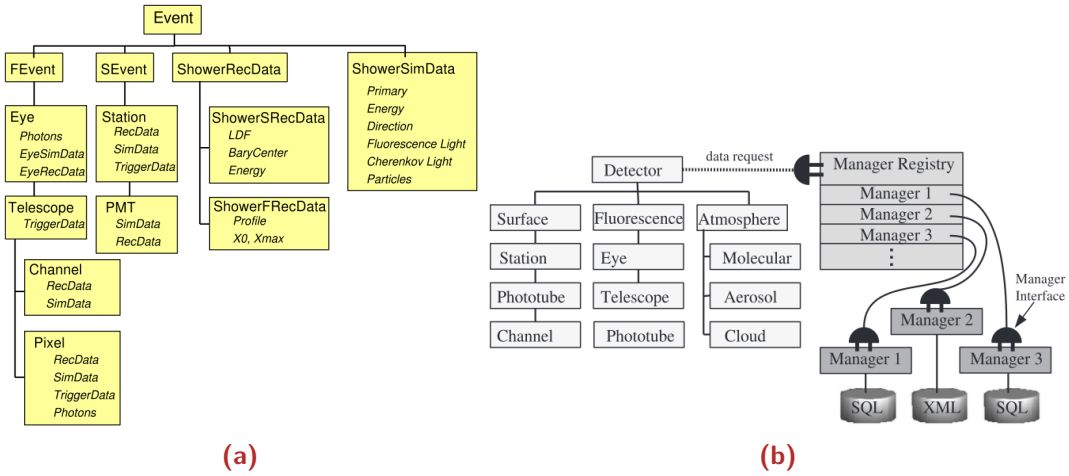


Figure 3.18. (a) Hierarchy of the Event structure and (b) machinery of the Detector description of the Auger Offline software framework [163].

3.4.4 Detector Description

In contrast to the event structure, the detector description acts as a read only data hierarchy and provides information about the state of the detector, its geometry, calibration and atmospheric conditions. For example, this includes the location of all SD stations, as well as calibration constants. This hierarchy is shown in Figure 3.18b.

```
<sequenceFile>
  <loop numTimes="unbounded">
    <module> SimulatedShowerReader </module>
    <loop numTimes="10" pushToStack="yes">
      <module> EventGenerator   </module>
      <module> TankSimulator    </module>
      <module> TriggerSimulator </module>
      <module> EventExporter    </module>
    </loop>
  </loop>
</sequenceFile>
```

Code 3.1 A simplified example of a Module Sequence set in an XML file for a simulation of the SD array. The <loop> command allows specific sets of tasks to be executed multiple times, for example they can be repeated for each event. The <module> command tells the run controller which module to execute, and the ordering of the file controls the sequence. In this module sequence, simulated showers are read from a file by the `SimulatedShowerReader` module, and each of these events are re-generated 10 times by `EventGenerator`. The two subsequent modules simulate the response of the detector and its trigger. Finally, `EventExporter` writes the data to disk. The pushToStack="yes" attribute stores the event at the beginning of the loop and restores it to its original state for the next loop so that the next random position can be fed through the simulated detector [163, 164].

3.5 Event Reconstruction

The reconstruction of measurements made by the SD and FD are processed offline utilising standard module sequences within the Offline framework. This includes geometrical reconstruction as well as reconstruction of the energy deposited in the atmosphere and the depth of shower maximum. The following section will outline the general procedures and algorithms used to reconstruct SD, FD and hybrid events.

3.5.1 Surface Detector Reconstruction

The reconstruction of the arrival direction and energy of an EAS which has triggered the SD array is based on the sizes of signals produced in individual SD stations, as well as the times at which those signals were registered. Each of the reconstruction steps are discussed in detail in the following sections.

3.5.1.1 Shower Geometry

A first guess of the shower geometry is estimated by fitting the trigger times of stations which satisfy the T4 trigger to a plane shower front moving at the speed of light (represented schematically in Figure 3.19) [136, 166]. For the first estimate of the time t_i that the shower front passes station i at position $\vec{x}_i$, this can be expressed as

$$-\hat{a}(\vec{x}_i - \vec{x}_b) = c(t_i - t_b) \tag{3.2}$$

where the vector $\hat{a}$ defines the shower axis (as defined in Figure 3.19), $\vec{x}_b$ defines the signal-weighted barycentre of the stations, with weighted bary-time t_b. The barycentre and bary-time are given by

$$\vec{x}_b = \frac{\sum_i \sqrt{S_i}\vec{x}_i}{\sum_i \sqrt{S_i}} \tag{3.3}$$

$$t_b = \frac{\sum_i \sqrt{S_i}t_i}{\sum_i \sqrt{S_i}} \tag{3.4}$$

where S_i is the total signal measured by station i. Based on the difference between the predicted time of arrival and the actual time of arrival at each station, the first estimate of the shower axis can be determined by minimising the following χ^2 function

$$\chi^2 = \sum_i \frac{\left(c(t_i - t_b) + \hat{a}\cdot(\vec{x}_i - \vec{x}_b)\right)^2}{\sigma_i^2} \tag{3.5}$$

where σ_i is the uncertainty in the start time of the signal in station i.

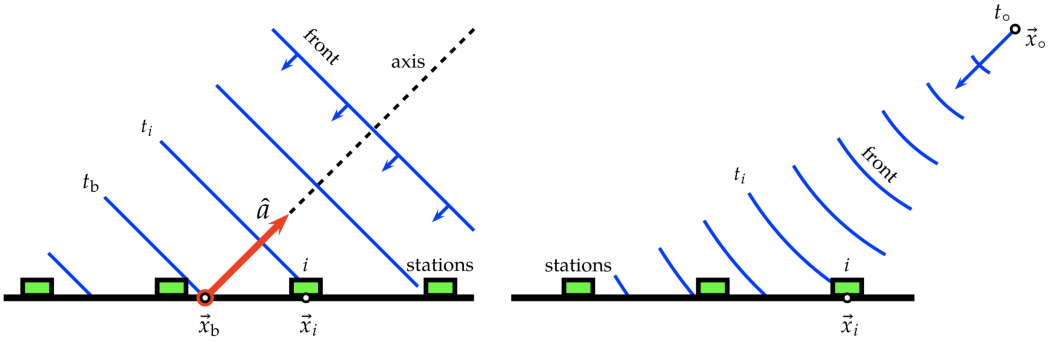


Figure 3.19. A schematic representation of the shower front evolution for a plane front approximation (left) and for a spherical shower front (right). Note some stations are triggered earlier than others; in this example the shower front reaches station i first at time t_i. This timing information allows the geometry of the shower to be determined by minimisation of the difference with respect to the expected time of arrival for a given shower axis [136].

In the next step, the model of the shower front is extended to a more realistic spherical shower front (represented schematically in Figure 3.19) [136, 166]. This can be expressed as

$$c(t_i - t_0) = |\vec{x}_0 - \vec{x}_i| \tag{3.6}$$

where $\vec{x_0}$ represents a virtual origin of the shower development at time t_0 [136]. The radius of curvature of the shower front R_c is determined during the reconstruction of the geometry similarly to the plane-front case by minimising the following χ^2 function

$$\chi^2 = \sum_i \frac{\left[c\left(t_i - t_0\right) + |R_c\hat{a} - \vec{x}_i|\right]^2}{\sigma_i^2} \tag{3.7}$$

In the case of only a few stations in the event, there is insufficient information to reconstruct the shower front curvature, so a parametrised model based on events with large numbers of stations is used to fix R_c.

Finally, the shower axis can be calculated by

$$\hat{a} = \frac{\vec{x}_0 - \vec{x}_c}{|\vec{x}_0 - \vec{x}_c|} \tag{3.8}$$

where $\vec{x}_c$ is the shower core position as determined by the lateral distribution function (LDF) reconstruction (see Section 3.5.1.2) [136].

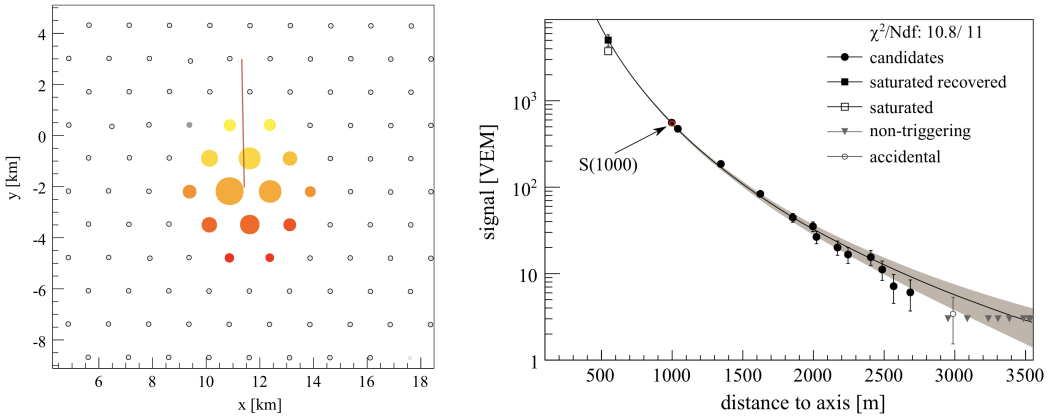


Figure 3.20. (a) A group of stations triggered by a shower. The coloured circles indicate tanks which have triggered, with the size of the circle being proportional to the logarithm of the signal. The colour gradient denotes the trigger timing; progressing in time from yellow to red. The shower axis (shown as the red line) can be fitted to determine the geometry of the shower [136] and (b) An example of the measured signal as a function of distance to the shower axis which is fit by the LDF using Equation (3.9) [136].

3.5.1.2 Lateral Distribution Function (LDF)

The lateral distribution of the SD signal can be described by a modified Nishimura-Kamata-Greisen (NKG) function [167, 168] where the signal in VEM, measured at a distance r from the shower axis, $S(r)$ is given by

$$S(r) = S(r_{\text{opt}}) \left(\frac{r}{r_{\text{opt}}} \right)^{\beta} \left(\frac{r + r_1}{r_{\text{opt}} + r_1} \right)^{\beta+\gamma} \tag{3.9}$$

where $r_1 = 700$ m, r_{opt} is the optimum distance (which depends on detector spacing) and β and γ are zenith angle dependent slope parameters [136]. Equation (3.9) is known as the lateral distribution function (LDF). For a given detector spacing, there exists an optimum shower-core distance, r_{opt} at which the dependence of the signal, $S\left(r_{\text{opt}}\right)$, on the parameter β is minimised [136]. Based on MC simulations, r_{opt} is found to be 1000 m for the detector spacing of 1500 m at Auger [169]. $S(r_{\text{opt}})$ is referred to hereafter as $S(1000)$. Similarly for the 750 m array r_{opt} is 450 m, with the signal denoted $S(450)$ [170]. An example of the footprint of a shower on the ground is shown in Figure 3.20a, and the lateral distribution of the signals is shown in Figure 3.20b along with the best fit of Equation (3.9).

The LDF fit is performed using a maximum likelihood technique which takes into account the different contributions from saturated, non-saturated and non-triggered stations [166]. The likelihood function is given by

$$\mathcal{L} = \prod_i f_P(n_i, \mu_i) \prod_i f_G(n_i, \mu_i) \prod_i f_{\text{sat}}(n_i, \mu_i) \prod_i f_{\text{zero}}(n_i, \mu_i) \tag{3.10}$$

where each component of the likelihood function represents the probability of observing a signal given the signal size and the station's state (either saturated or not). Given the total signal of each station, the effective particle number can be calculated based on the zenith angle of the shower. The effective particle number is given by

$$n_i = \frac{S_i}{l(\theta)} \tag{3.11}$$

where $l(\theta)$ is the mean relative track-length of through-going particles crossing the detector at zenith angle θ. The components of the likelihood function are described in detail below.

1. **Small Signals**
   In the case of small signals, the probability is determined by a Poisson distribution
   $$f_P(n_i, \mu_i) = \frac{\mu_i^{n_i} e^{-\mu_i}}{n_i!} \tag{3.12}$$
   where n_i is the effective particle number and μ_i is the expected signal from the LDF.

2. **Large Signals**
   Larger signals use the Gaussian approximation for the probability
   $$f_G(n_i, \mu_i) = \frac{1}{\sqrt{x\pi}\sigma_i} \exp\left(-\frac{(n_i - \mu_i)^2}{2\sigma_i^2}\right) \tag{3.13}$$
   where σ_i represents the uncertainty in the measured signal.

3. **Saturated Signals**
   In the case that a saturated signal can be recovered it is treated the same as in the Gaussian approximation for large signals with the uncertainty of the recovery added to σ_i. However, in the case of a saturated signal that cannot be recovered, an upper limit is determined by integrating over all of the possible values larger than the measured n_i, using the Gaussian approximation
   $$f_{\text{sat}}(n_i, \mu_i) = \int_{n_i}^{\infty} f_G(n, \mu_i) dn. \tag{3.14}$$

4. **Zero Signals**
   Stations which have not triggered may have done so if the signal was anywhere below the threshold to form a trigger. As such, all possible values of the signal up to the threshold are summed using the Poisson probability for small signals

$$f_{\text{zero}}(n_i, \mu_i) = \sum_{n_i=0}^{n_{\text{th}}} f_P(n_i, \mu_i) \tag{3.15}$$

3.5.1.3 Energy Reconstruction

Once the best fit LDF has been determined, an estimate of the shower energy can be obtained using $S(1000)$, the signal that would be measured at a distance of 1000 m from the shower axis. However for a given energy, $S(1000)$ has a dependence on zenith angle due to the increasing atmospheric attenuation for large zenith angles. To correct for this effect, the dependence of $S(1000)$ on zenith angle is removed by converting to the quantity S_{38}, which is a zenith-angle independent energy estimator and can be thought of as the $S(1000)$ signal produced if the shower had arrived at a zenith angle of $38°$; the median of the zenith angle distribution. Similarly for the 750 m array, $S(450)$ is converted to S_{35} [170]. S_{38} is derived from data using the Constant Intensity Cut (CIC) method, which relies on the (reasonable) assumption that the flux of cosmic rays at the top of the atmosphere is isotropic [171]. Bins of zenith angle are chosen to provide a uniform acceptance in solid angle. The attenuation curve $f_{\text{CIC}}(\theta)$ is fitted with a third-order polynomial of the form

$$f_{\text{CIC}}(\theta) = 1 + ax + bx^2 + cx^3 \tag{3.16}$$

where the constants $a = 0.980 \pm 0.004$, $b = -1.68 \pm 0.01$, $c = -1.30 \pm 0.45$, and $x = \cos^2\theta - \cos^2 38°$. Given $S(1000)$ and zenith angle θ, S_{38} can be calculated by

$$S_{38} = \frac{S(1000)}{f_{\text{CIC}}(\theta)} \tag{3.17}$$

Having derived a reference parameter independent of zenith angle, a calorimetric energy estimator can be obtained by a cross-calibration with the FD energy measurements (see Section 3.5.2.7). A strong correlation between S_{38} and E_{FD} is evident in Figure 3.21. S_{38} is calibrated to E_{FD} using the functional form

$$E_{\text{FD}} = A\left(\frac{S_{38}}{\text{VEM}}\right)^B \tag{3.18}$$

where $A = (0.186 \pm 0.003) \times 10^{18}$ eV and $B = 1.031 \pm 0.004$ are the best fit parameters derived from high quality "Golden" hybrid events which independently trigger the FD and SD [22].

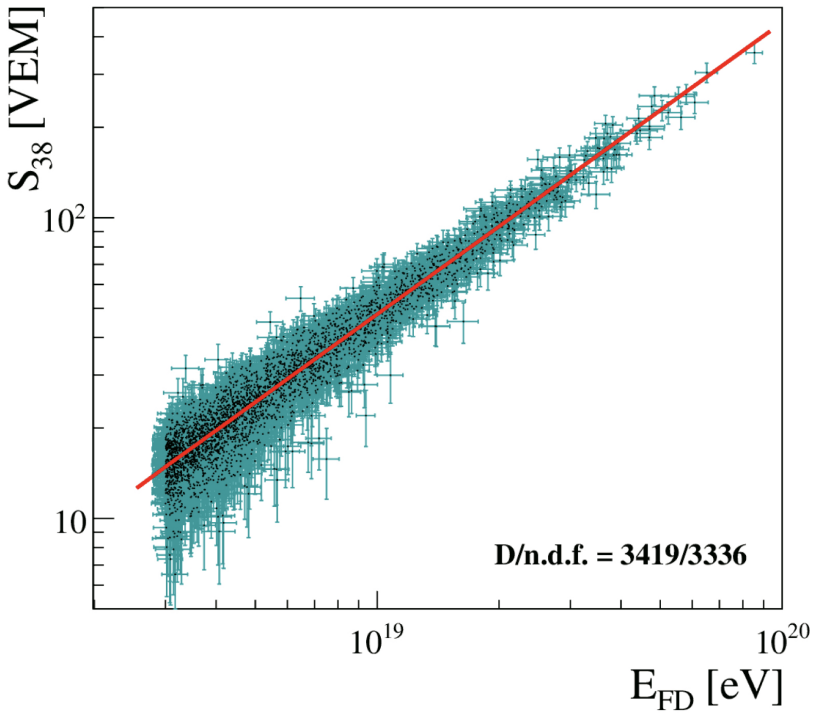


Figure 3.21. Correlation between the FD energy and the SD energy estimator S_{38}. The red line represents the best fit calibration for Equation (3.18) [22].

3.5.2 Fluorescence Detector Reconstruction

The reconstruction procedure is divided into several steps, each controlled by a separate Offline module. The fluorescence light emitted by air showers is detected as a series of triggered pixels across the camera of the FD. An example of this is shown in Figure 3.22. The orientation of the shower axis is determined using the pixel pointing directions and the timing sequence of the pulses in the pixels of the camera. The signals in these pulses are used to determine the energy deposited in the atmosphere. Each of the reconstruction steps are discussed in detail in the following sections.

3.5.2.1 Night-Sky Background

Each pixel measures a slightly different level of background noise. Since the PMT electronics are AC-coupled, it is not possible to directly measure the level of night-sky background (NSB). It is instead estimated from the variance of the ADC signals in early time bins of the pixel's trace (which are free from any shower signal). The

mean of the signal is subtracted from the entire pixel trace. The total measured signal variance is the combination of the NSB and the electronic noise

$$\sigma^2 = \sigma^2_{\text{NSB}} + \sigma^2_{\text{electronic}}. \tag{3.19}$$

The variance is converted from ADC counts to photoelectrons by dividing by the square of the absolute gain, A_{G}

$$\sigma^2_{\text{pe}} = \frac{\sigma^2_{\text{ADC}}}{A^2_{\text{G}}} \tag{3.20}$$

where

$$A^2_{\text{G}} = \eta G E = \frac{1}{C_{\text{FD}} f Q} \tag{3.21}$$

and η is the PMT collection efficiency at the first dynode, G is the PMT gain, E is the electronics conversion constant (units of ADC counts per electron), C_{FD} is the pixel calibration constant, Q is the PMT quantum efficiency and f is the telescope optical factor which depends on filter transmission, mirror reflectivity and collection efficiency.

Using this information, the NSB variance can be converted into a signal in photoelectrons by

$$n_{\text{pe}} = \frac{\sigma^2_{\text{pe}}}{1 + V_g} \tag{3.22}$$

where V_g is the gain variance of the PMT. This can then be converted into a photon flux by

$$\Phi_\gamma = \frac{n_{\text{pe}}}{Q f \mathcal{A} \Delta t} \tag{3.23}$$

where f is the telescope optical factor, $\mathcal{A}$ is the pixel aperture and Δt is the sampling bin size of 100 ns (50 ns for HEAT).

3.5.2.2 Pixel Pulse Search

Each FD pixel trace is searched to locate a pulse start and stop time corresponding to a shower signal. These times are determined using an algorithm implemented by the `FdPulseFinder` module. The background noise σ_i for the ith pixel is estimated from early bins which are free from any shower signal. The total signal in the ith pixel within a sliding "boxcar" window is given by

$$S_i = \sum_{k=k_{\text{start}}}^{k_{\text{stop}}} S_{ik} \tag{3.24}$$

where the range k_{start} and k_{stop} is chosen to maximise the signal-to-noise ratio which is defined as

$$\frac{S}{N} = \frac{S_i}{\sigma_i \sqrt{k_{\text{stop}} - k_{\text{start}}}}. \tag{3.25}$$

A signal-to-noise ratio ≥ 5 is required for a pixel to be considered in the reconstruction [136]. The module also determines the ith pixel's centroid time (signal weighted time average of the pulse) which is given by

$$t_i = \frac{\sum \tau_k^i s_k^i}{\sum s_k^i} \tag{3.26}$$

where τ_k and s_k are the time and signal of the kth pulse bin.

The centroid pulse time uncertainty is determined by the propagation of errors from the noise variance and photo-electron signal fluctuations through Equation (3.26) [136].

3.5.2.3 Shower Detector Plane

The geometry of the shower axis is determined using a two-stage algorithm. The first stage is the reconstruction of the shower-detector plane (SDP). The SDP is defined as the plane which contains the location of the FD and the shower axis and is illustrated in Figure 3.23. This concept is useful because (by definition) the line formed by the projection of the shower onto the FD camera is the edge of the SDP. The SDP is reconstructed from the data by minimising

$$S = \frac{\sum_i q_i \left(\frac{\pi}{2} - \arccos\left(\vec{p_i} \cdot \vec{n}_{\perp}^{\text{SDP}} \right) / \sigma_{\text{SDP}} \right)^2}{\sum_i q_i} \tag{3.27}$$

over all pulses i. Here $\vec{p_i}$ is the ith pixel pointing direction, q_i is the integral of the pulse signal, $\vec{n}_{\perp}^{\text{SDP}}$ is defined as the normal vector to the SDP (defined by the free fit parameters θ_{SDP} and ϕ_{SDP}) and σ_{SDP} was determined empirically to be 0.35° [136]. The line through the triggered pixels in Figure 3.22 shows an example of the fitted SDP.

3.5.2.4 Reconstruction of the Shower Axis

The second stage of the geometry reconstruction is the determination of the shower axis within the SDP. Pixel timing information is used in the `FdAxisFinder` module to determine the orientation of shower axis within the SDP utilising the shower parameters shown in Figure 3.23, and fitting to the following functional form

$$t_i = t_0 + \frac{R_p}{c} \tan\left(\frac{\chi_0 - \chi_i}{2} \right) \tag{3.28}$$

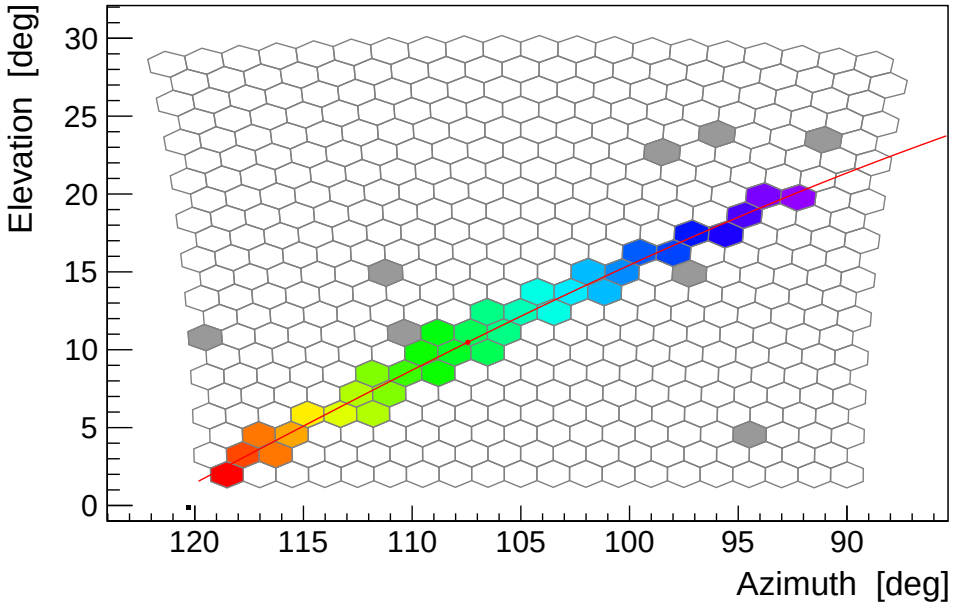


Figure 3.22. An example event as it appears in the FD camera. The pattern of triggered pixels shows the development of the shower and the colours indicate the timing sequence (progressing in time from purple to red). The red line is the best-fit SDP projected on the FD camera. The grey pixels are those that were triggered but did not pass the reconstruction algorithm because they were too far in angle or in time to the best-fit SDP.

where t_i is the time that the signal arrives at the ith pixel and t_0 is defined as the time when the shower front passes the point of closest approach to the camera, R_p [144]. The other parameters are defined in Figure 3.23. To determine the three free fit parameters R_p, t_0 and χ_0, the following χ^2 is minimised.

$$\chi^2 = \sum_i \frac{(t_i - t(\chi_i))^2}{\sigma(t_i)^2} \tag{3.29}$$

Unfortunately, when only short tracks are observed by the FD, R_p and χ_0 are degenerate due to the functional form of Equation (3.28). This introduces uncertainties in other shower parameters such as energy and $X_{\max}$. Additional timing information from the SD (available only for hybrid events which also trigger SD stations) can be added to Equation (3.29) in order to break the degeneracy between parameters. Hybrid reconstruction is discussed in Section 3.5.3.

3.5.2.5 Light at the Aperture

The `FdApertureLightFinder` module is used to calculate the total photon flux received at the aperture of the FD telescopes as a function of time. Firstly, for each 100 ns time bin, signals from the camera pixels are added together. Pixels are only included if their pointing direction is within an angle ζ of the reconstructed shower

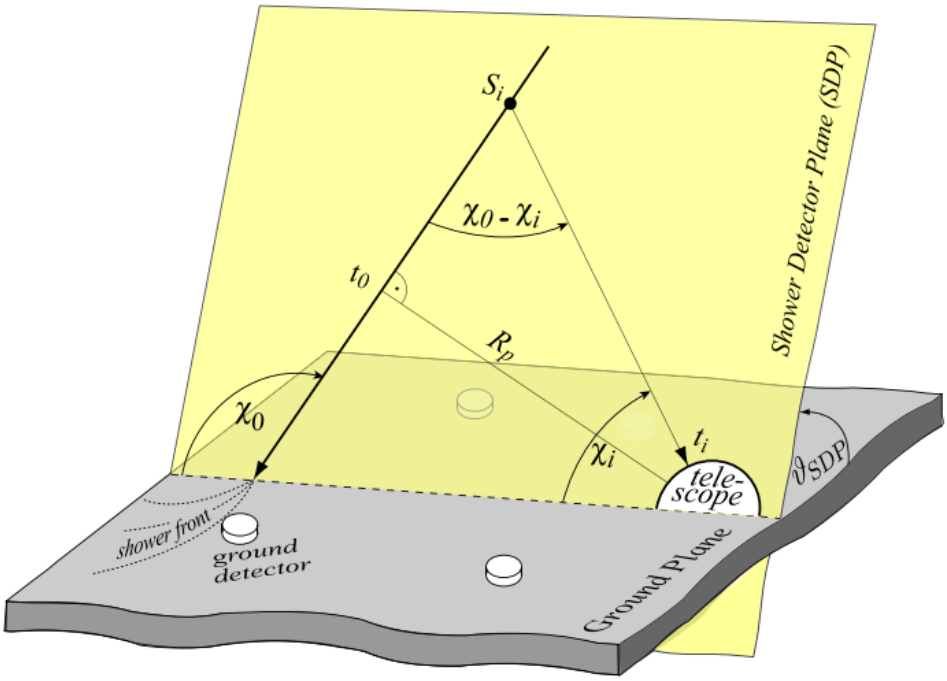


Figure 3.23. An illustration of the SDP and the parameters of the shower axis timing fit [144].

position in a given time bin. The value of ζ is chosen such that the signal-to-noise ratio is maximised. The total light flux at the aperture in time bin i is

$$F_i = \frac{1}{A_{\text{dia}}} \sum_{j=1}^{N_{\text{pix}}} S_{ij} \tag{3.30}$$

where A_{dia} is the area of the diaphragm opening, N_{pix} is the number of pixels and j runs only over pixels within ζ at time t_i.

3.5.2.6 Shower Profile

Once the geometry of the shower has been reconstructed, the photon flux measured by the FD as a function of time must be converted to the energy deposited by the shower in the atmosphere as a function of slant depth. The shower profiles before and after this conversion are shown in Figures 3.24 and 3.25, respectively.

In order to calculate the energy deposited as a function of slant depth (atmospheric depth), models of the fluorescence and Cherenkov yields are required [172]. The

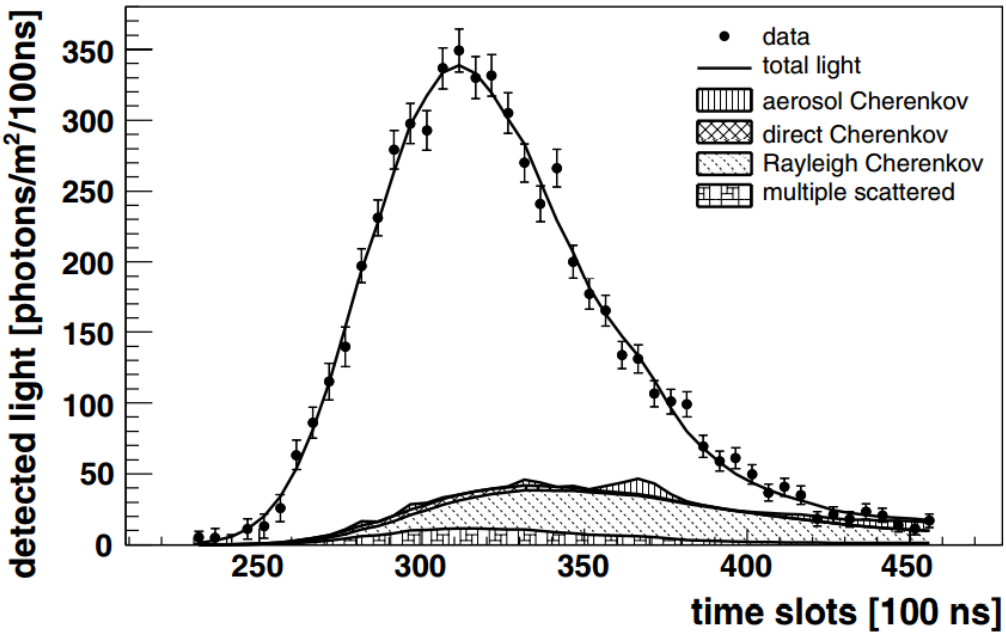


Figure 3.24. The flux of the sources of light detectable by the FD cameras as a function of time. The fluorescence light can be identified as the difference between the total detected light, and the contributions from Cherenkov and multiple scattered photons. These sources of light must be accounted for before converting to energy deposited as a function of atmospheric depth, shown in Figure 3.25[144].

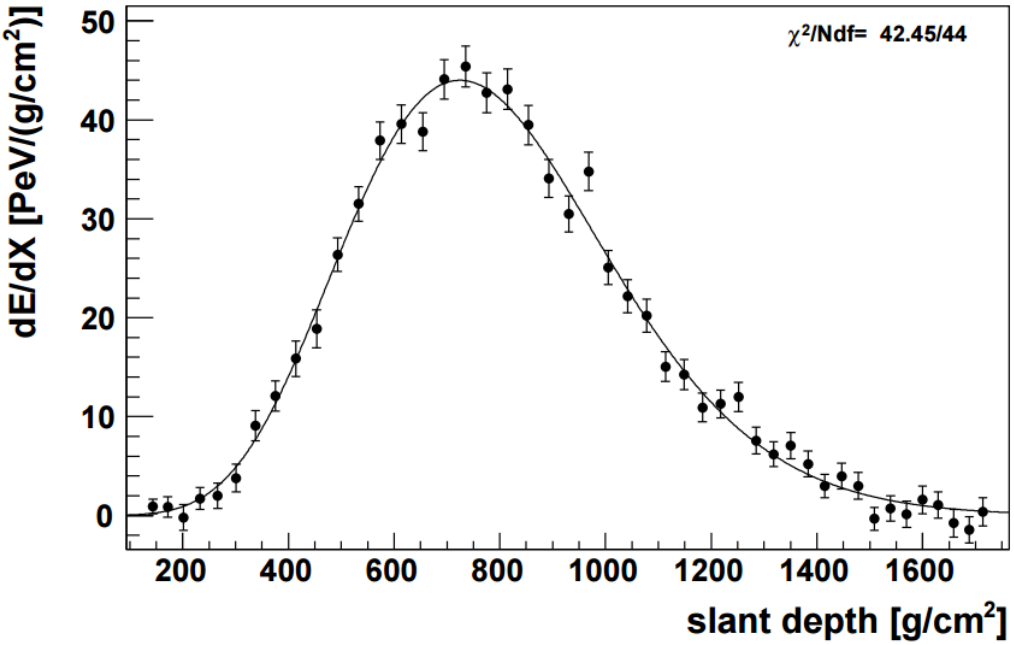


Figure 3.25. The energy deposited by the air shower as a function of slant depth with a fitted Gaisser-Hillas function. This curve is integrated to give the calorimetric energy deposited by the shower. Note the maximum of the Gaisser-Hillas function is at a depth $X_{\max}$ [144].

number of fluorescence photons produced by the shower at a depth X_i can be written as

$$N_\gamma^{\mathrm{f}}(X_i) = \Delta X_i \frac{dE}{dX_i} \sum_k Y_{ik}^{\mathrm{f}} \tag{3.31}$$

where Y_{ik}^{f} is the fluorescence yield at wavelength λ_k and dE/dX_i is the energy deposited in a small range of atmospheric depth ΔX_i. However, not all photons reach the detector. Since the photons are emitted isotropically, only a small fraction will fall within the solid angle subtended by the detector FoV. This reduces the photon flux measured by a factor of $1/r^2$. Along the path from the shower to the detector, only a fraction T_{ik} of photons will survive being scattered away from the detector. Finally, the photon flux is reduced by the detector efficiency, ϵ_k. The fluorescence light received at the detector as a function of depth can therefore by written as

$$y_i^{\mathrm{f}} = \Delta X_i \frac{dE}{dX_i} \sum_k \frac{\epsilon_k T_{ik}}{4\pi r_i^2} Y_{ik}^{\mathrm{f}} \tag{3.32}$$

Similarly for Cherenkov light, the number of photons produced can be written as

$$N_\gamma^{\mathrm{C}}(X_i) = \Delta X_i N_i^e \sum_k Y_{ik}^{\mathrm{C}} \tag{3.33}$$

where the production depends on N_i^e, the number of charged particles above the Cherenkov threshold energy, and the Cherenkov yield Y_{ik}^{C}. Unlike fluorescence light, Cherenkov light is forward beamed so only a fraction $f_d(\beta i)$ of the photons originate directly from the shower. The fraction detected depends on the angle of observation with respect to the shower axis, β_i. The direct Cherenkov light received at the detector can therefore by written as

$$y_i^{\mathrm{Cd}} = \Delta X_i N_e(X_i) f_{\mathrm{d}}(\beta_i) \sum_k \frac{\epsilon_k T_{ik}}{4\pi r_i^2} Y_{ik}^{\mathrm{C}} \tag{3.34}$$

This can be rewritten in a more useful form using the mean ionisation loss rate $\alpha(X_i)$, which is the average energy deposited per unit depth per electron at shower age $s_i = 3/(1 + 2X_{\mathrm{max}}/X_i)$

$$y_i^{\mathrm{Cd}} = \frac{1}{\alpha_{\mathrm{i}}} \frac{dE}{dX_i} \Delta X_i f_{\mathrm{d}}(\beta_i) \sum_k \frac{\epsilon_k T_{ik}}{4\pi r_i^2} Y_{ik}^{\mathrm{C}} \tag{3.35}$$

Parametrisations of α can be found in [173, 174]. Due to the strong forward beaming of Cherenkov light production, an intense beam of Cherenkov photons builds up along the shower axis. The number of photons in the Cherenkov beam at depth X_i is the combined number of all photons generated at all previous depths X_j, reduced only by their attenuation T_{jik}

$$N_\gamma^{\mathrm{beam}}(X_i) = \sum_{j=0}^{i} \Delta X_j N_j^e \sum_k \tau_{jik} Y_{jk}^{\mathrm{C}} \tag{3.36}$$

However, only the fraction $f_s(\beta_i)$ of the photons which scatter out of the beam eventually reach the detector. This depends on the scattering angle β_i. The scattered Cherenkov light received at the detector can therefore by written as

$$y_i^{\mathrm{Cs}} = f_{\mathrm{s}}\left(\beta_i\right) \sum_{j=0}^{i} \Delta X_j N_j^e \sum_k \frac{\epsilon_k T_{ik}}{4\pi r_i^2} \tau_{jik} Y_{jk}^{\mathrm{C}} \tag{3.37}$$

Similarly to the case of direct Cherenkov light, Equation (3.37) can be rewritten using the mean ionisation loss rate α to give

$$y_i^{\mathrm{Cs}} = f_{\mathrm{s}}\left(\beta_i\right) \sum_{j=0}^{i} \Delta X_j \frac{1}{\alpha_j} \frac{dE}{dX_j} \sum_k \frac{\epsilon_k T_{ik}}{4\pi r_i^2} \tau_{jik} Y_{jk}^{\mathrm{C}} \tag{3.38}$$

Therefore, the total light received at the detector is given by

$$y_i = y_i^{\mathrm{f}} + y_i^{\mathrm{Cd}} + y_i^{\mathrm{Cs}} \tag{3.39}$$

Clearly Equation (3.39) can now be used to calculate dE/dX_i as a function of X_i, which is the shower profile. It is convenient to convert Equation (3.39) into matrix form. Let $\mathbf{y} = (y_1, \ldots y_n)^T$ and $\mathbf{x} = (dE/dX_1, \ldots dE/dx_n)^T$. Then Equation (3.39) can be written in the form

$$\mathbf{y} = \mathbf{C}\mathbf{x} \tag{3.40}$$

where $\mathbf{C}$ is called the Cherenkov-fluorescence matrix and is given by

$$C_{ij} = \begin{cases} 0 & i < j \\ c_i^{\mathrm{d}} + c_{ii}^{\mathrm{s}} & i = j \\ c_{ij}^{\mathrm{s}} & i > j \end{cases} \tag{3.41}$$

where

$$c_i^d = \Delta X_i \sum_k \frac{\epsilon_k T_{ik}}{4\pi r_i^2} \left(\frac{f_d(\beta_i)}{\alpha_i} Y_{ik}^{\mathrm{C}} + Y_{ik}^{\mathrm{f}} \right) \tag{3.42}$$

and

$$c_{ij}^s = \Delta X_j \sum_k \frac{\epsilon_k T_{ik}}{4\pi r_i^2} \tau_{jik} \frac{f_{\mathrm{s}}\left(\beta_i\right)}{\alpha_{\mathrm{j}}} Y_{jk}^{\mathrm{C}} \tag{3.43}$$

The solution to Equation (3.39) can then be determined by matrix inversion

$$\mathbf{x} = \mathbf{C}^{-1}\mathbf{y} \tag{3.44}$$

Since the elements of $\mathbf{C} \geq 0$, $\mathbf{C}$ is always invertible. The statistical uncertainties of $\mathbf{x}$ are obtained by error propagation

$$\mathbf{V}_x = \mathbf{C}^{-1} \cdot \mathbf{V}_y \cdot \left(\mathbf{C}^{\mathrm{T}}\right)^{-1} \tag{3.45}$$

where $\mathbf{V}_y$ is the uncertainty in the measured light flux and is dominated by photo-electron statistics at the PMT level [172].

3.5.2.7 Gaisser-Hillas Profile Fit

In order to determine the calorimetric energy from the reconstructed shower profile, the undetected parts of the shower must be extrapolated from the measured parts. This is done through a maximum likelihood fit of the shower profile to a Gaisser-Hillas function [175] of the form

$$f_{\mathrm{GH}}(X) = \left(\frac{dE}{dX}\right)_{\mathrm{max}} \left(\frac{X - X_0}{X_{\mathrm{max}} - X_0}\right)^{(X_{\mathrm{max}} - X_0)/\lambda} \exp\left(\frac{X_{\mathrm{max}} - X}{\lambda}\right) \tag{3.46}$$

In the case where a large fraction of the shower is outside the FoV then all four Gaisser-Hillas parameters cannot be reconstructed. In this scenario, the shape parameters X_0 and λ are constrained to a range of typical values $\langle X_0 \rangle = -121 \pm 172$ g/cm^2 and $\langle \lambda \rangle = 61 \pm 13$ g/cm^2, respectively. The calorimetric energy of the shower is then given by the integral of the best-fit Gaisser-Hillas function

$$E_{\mathrm{cal}} = \int_0^\infty f_{\mathrm{GH}}(X) dX. \tag{3.47}$$

The total energy of the shower is estimated using a correction for the "invisible energy" carried away by neutrinos and high energy muons, which is not deposited in the atmosphere and represented as fluorescence light [172]. The invisible energy is estimated using a data-driven approach, which is based on measurements from the SD in hybrid events (Section 3.5.3). The correlation between the invisible energy E_{inv}, and the calorimetric energy E_{cal}, can then be approximated by a power law

$$E_{\mathrm{inv}} = a \left(\frac{E_{\mathrm{cal}}}{\mathrm{EeV}}\right)^b \tag{3.48}$$

where $a = 0.179 \pm 0.006$ and $b = 0.947 \pm 0.017$ for inclined events and $a = 0.160 \pm 0.002$ and $b = 0.952 \pm 0.005$ for vertical events [176].

Recent modifications to the Auger reconstruction procedure in 2019 sees the introduction of a new Gaisser-Hillas function for the shower profile, modified from the original to use two new shape parameters R and L [177]. The new form of the Gaisser-Hillas function is

$$f(X) = \left(\frac{dE}{dX}\right)_{\mathrm{max}} \left(1 + \frac{R}{L}(X - X_{\mathrm{max}})\right)^{1/R^2} \exp\left(-\frac{X - X_{\mathrm{max}}}{RL}\right) \tag{3.49}$$

where L is a measure of the width of the shower profile, and R is an asymmetry parameter. These new parameters are not correlated like the two parameters, X_0 and λ that they replace [177, 178].

3.5.3 Hybrid Event Reconstruction

The reconstruction of hybrid events is based on FD measurements, along with additional timing information provided by SD stations. The resolution of FD monocular reconstruction can be limited for many events, however since the SD operates continuously, most of the events detected by the FD are also detected by at least one SD station. These events are known as hybrid events.

The information from just a single station can suffice for the hybrid reconstruction process. This is because the geometry reconstruction is vastly improved by using the arrival time of the shower at the ground, relatively far away from the bulk of the shower track observed by the FD. This is especially useful for short angular track lengths seen by the FD cameras, as mentioned in Section 3.5.2. Equation (3.50) shows how the additional information is used to modify Equation (3.29) and constrain the shower geometry.

$$\chi^2 = \sum_i \left(\frac{\left(t_i - t\left(\chi_i\right)\right)^2}{\sigma\left(t_i\right)^2} + \frac{\left(t_{\mathrm{SD}} - t\left(\chi_{\mathrm{SD}}\right)\right)^2}{\sigma\left(t_{\mathrm{SD}}\right)^2} \right) \tag{3.50}$$

Figure 3.26 shows an example of the improvement in the uncertainties of the reconstructed shower geometry (R_p, χ_0 and T_0) when utilising the additional information provided by the SD in comparison to the standard FD monocular reconstruction. This more accurate determination of the shower geometry using a hybrid combination of detectors ensures that the energy and depth of shower maximum is resolved as accurately and precisely as possible.

3.6 AugerPrime Upgrade

Auger is currently undergoing an upgrade called “AugerPrime” in order to improve measurements of the flux suppression (Section 2.2.3) and mass composition at the highest energies. [179–181]. AugerPrime consists of the following components:

- **Surface Scintillation Detectors (SSDs)**
  Each of the SD stations will be upgraded with a plastic scintillator added to the roof of the water tank in order to provide a complementary measurement of shower particles. The existing Water Cherenkov Detector (WCD) and the

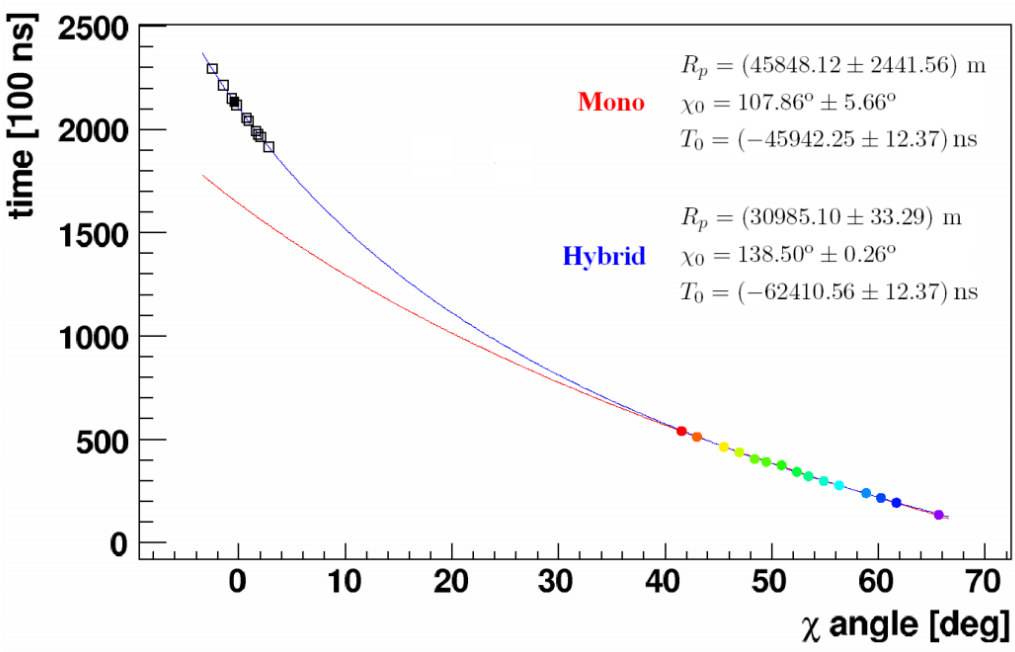


Figure 3.26. An example of the improved reconstruction that the hybrid measurements provide. The FD data (coloured points) and SD data (squares) are superimposed. Only a single SD station is required for hybrid reconstruction [144].

new SSD will provide a different response to the muonic and electromagnetic components of the shower. This additional information will contribute to an event-by-event measurement of the mass composition. A photograph of an upgraded station is shown in Figure 3.27. The SSD units house two scintillator sub-modules, providing a total collecting area of 3.8 m^2 per detector [179, 180, 182]. A pre-production array of 77 SSD units have been deployed and have begun data acquisition [181].

- **Underground Muon Detector (UMD)**
  A component of the AMIGA enhancement to Auger [183] was an array of underground muon detectors in the 750 m array. Each detector consists of 3 scintillators providing a total area of 30 m^2. They are buried 2.3 m underground to attenuate the electromagnetic component of showers. This provides a direct measurement of the muon content of showers. The UMD will consist of 61 AMIGA muon detectors and is expected to be completed by the end of 2020 [179, 180, 184].

- **New SD Electronics**
  The current SD electronics have become outdated since their design over 20 years ago. Improvements in technology have been used to design and begin implementing a new, higher performance electronics system (120 MHz sampling compared to 40 MHz) for the SD array and the SSD upgrade. In

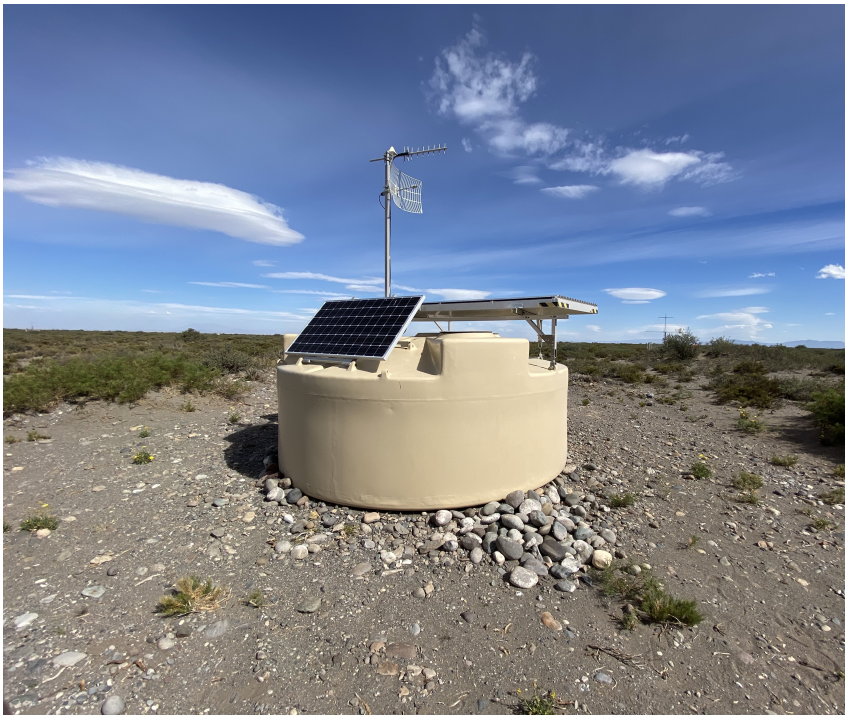


Figure 3.27. An upgraded SD station with a SSD added to the roof of the tank. Photograph taken by the Author.

addition to the new electronics system, an additional small PMT will be added to the WCDs in order to increase their dynamic range. This is expected to reduce the number of saturated stations by a factor of 10 at the highest energies [179, 180, 185, 186].

- **Extended FD Operation**
  The FD duty cycle is currently limited by the requirement of moonless nights for operation. The criteria for FD measurements to be recorded on a given night are that the Sun must be 18° below the horizon, and the Moon must nominally be below the horizon for 3 hours. These requirements limit the FD operation time to $\sim$15%, however an increase up to a 29% duty cycle could be achieved by relaxing the requirement of the Moon being below the horizon. In order to preserve the lifetime and stability of the FD PMTs, the supplied HV can be reduced by a factor of 10. Such an improvement in the FD duty cycle would provide up to 40% more events above $10^{19.5}$ eV. On-site and laboratory tests at a reduced HV have provided encouraging results [179, 180].

- **Large Radio Detector**
  A prototype array called Auger Engineering Radio Array (AERA) has been operating at Auger since 2011. Currently the array comprises 153 radio stations, covering an area of $\sim$17 km^2 [187, 188]. The measurement technique relies on the radio emission produced by the interaction of the air shower with

the Earth's magnetic field. For vertical showers, the radio footprint is very small on the ground, however very inclined showers produce a footprint of several km^2, making them ideal candidates for a sparse radio array. As part of the AugerPrime upgrade, a radio antenna will be installed on each of the SD stations. Upon completion of the upgrade, the array will become the largest cosmic ray radio detector in the world [189].

4 Energy Threshold for Full Trigger Efficiency

4.1 Introduction

In this chapter, the possibility of reducing the energy threshold for the HEAT-Coihueco (HeCo) X_{max} analysis using the new Surface Detector (SD) triggers in hybrid events is investigated. These events trigger both the High-elevation Auger telescopes (HEAT) and Coihueco along with the 750 m array. It is critical that there is no composition bias present in the SD trigger for such analyses. Therefore, a data driven approach using hybrid events is adopted in order to avoid extrapolation of hadronic models for simulations. Real hybrid data contains the true mass composition at a given energy and will therefore provide a more reliable estimate of the trigger efficiency. Parametrisations of the Lateral Trigger Probability (LTP) are derived from hybrid data rather than from detector simulations. In Section 4.3, a toy Monte Carlo (MC) simulation, is used to evaluate the array trigger efficiency as a function of energy and zenith angle for several trigger conditions and array spacings using the derived LTP parametrisations for single stations.

In 2013, two new SD triggers (time-over-threshold de-convoluted (ToTd) and multiplicity of positive steps (MoPS)) were introduced in order to lower the energy threshold of the SD [136]. The trigger efficiency of both the Fluorescence Detector (FD) and SD have been thoroughly investigated in previous simulation studies of the detector response to the original time-over-threshold (ToT) and threshold (Th) triggers (referred to as "old" triggers) as well as the new triggers. The analyses presented in this chapter are focused on the energy threshold for full efficiency of the new triggers for a single station (i.e. the requirement for a hybrid event). This is most relevant to X_{max} analysis which currently includes data down to $10^{17.2}$ eV and $10^{17.8}$ eV for the 750 m and 1500 m arrays, respectively [93, 97]. The SD T4 trigger efficiency is also studied in detail and compared with previous simulation studies in order to validate the results.

A secondary achievement of this study is a new, computationally fast procedure for estimating the trigger efficiency for any configuration of the SD. Since this method uses hybrid data to derive parametrisations of the single station trigger probability, it

does not require computationally expensive simulations or subsequent reconstruction procedures in order to generate the data set. Once the LTP has been derived, the trigger efficiency of any configuration of SD stations can be simulated very quickly. This makes the entire procedure very efficient.

4.2 Trigger Probability for a Single Station

4.2.1 Lateral Trigger Probability

In order to understand the trigger efficiency of the combination of different detectors, it is critical to first understand the trigger probability of a single detector. The probability for a single station depends on several independent parameters which can be formalised by the LTP: $LTP_{A,E,\theta,Tr}(r)$. The LTP is the probability of an air shower induced by a primary particle of energy E, mass A and zenith angle θ triggering a station [190]. Here, r is the distance between the triggered station and the shower axis in the shower plane (station-axis distance). There is no azimuthal angle dependence considered in this analysis given that the results from simulations show any such dependence does not introduce measurable differences [190].

Given a particular trigger condition Tr, the probability is defined as:

$$LTP_{A,E,\theta,Tr}(r) = \frac{N_1}{N_1 + N_0} \tag{4.1}$$

where N_1 and N_0 are the number of triggered and un-triggered stations, respectively, at a distance r from the shower axis.

The LTP functions and their parametrisations are derived in this chapter using hybrid data. Although simulated LTP parametrisations generally show good agreement with data [190], data derived LTP parametrisations avoid the need for extrapolation of hadronic models since they naturally contain the true mass composition information of the events. Having said this, it is important to consider the differences which do exist between data and simulation and why they exist. This is discussed in Section 4.2.7.

The data set used in this analysis is from 2014 (when the new triggers were implemented) onwards. All FD monocular events are removed to ensure a good geometry reconstruction. This is important due to the strong dependence of the geometry on the LTP. Additionally, the hybrid station (i.e. the station used for the hybrid geometry fit) is always ignored in order to limit the bias of events with only a single station, which can inherently produce an LTP of 1. Independent LTP parametrisations for

each of the possible trigger combinations for hybrid and SD-T4 triggers are also derived. This is done because hybrid events have different requirements compared with SD events. More specifically, hybrid events are formed when an FD trigger and any SD stations have triggered in time coincidence. However, for an SD event to trigger independently of the FD it must form either a 3ToT or 4C1 T4 trigger, as discussed in Section 3.1.2. In order to study both single station and SD T4 trigger efficiencies, as well as to compare old and new single station triggers there, are six different LTP parametrisations derived:

1. ToT (old triggers for 3ToT SD events)

2. ToT || TH-T1 || TH-T2 (old triggers for hybrid events)

3. ToT || TH-T1 || TH-T2 || ToTd || MoPS (new triggers for hybrid events)

4. ToT || ToTd || MoPS (new triggers for 3ToT SD events)

5. ToT || TH-T2 (old triggers for 4C1 SD events)

6. ToT || TH-T2 || ToTd || MoPS (new triggers for 4C1 SD events)

For each hybrid event, stations are binned in 100 m radius intervals from the shower axis. For each radius bin the number of triggered stations and the total number of working stations are counted. The LTP is then determined by the ratio described in Equation (4.1). This procedure is repeated for all hybrid events and an average LTP as a function of station-axis distance is calculated for each of the trigger configurations as well as for a variety of energy and zenith angle bins. The data are then fitted using Equation (4.7) as described in Section 4.2.2.

4.2.2 Fit Function

In previous simulation studies of the LTP, several variations of the LTP fit function have been used. In general, a fit combining a step function (close to the shower axis) with an exponential (further away) is found to reproduce the simulated data sets reasonably well [190]. The explicit form of the fit function is shown in Equation (4.2).

$$LTP(r) = \begin{cases} \frac{1}{1+e^{\frac{r-R_0}{\Delta R}}} & \text{for } r \leq R_0 \\ \\ \frac{1}{2}e^{A(r-R_0)} & \text{for } r > R_0 \end{cases} \tag{4.2}$$

where R_0, ΔR and A are free fit parameters. R_0 is the distance where the step function equals 0.5, ΔR and A control how quickly the step function, and the exponential, respectively, fall to zero. This form of the fit function was used in the 2011 published paper [190].

Upon consideration of the correlation between the fit parameters (R_0, ΔR and A) and uncertainty propagation of the subsequent analysis, it was discovered that the two parameters ΔR and A are related. In fact, the correlation is so strong that the fits do significantly better with only two parameters. To understand why, consider the continuity of the fit function and its derivatives. Firstly, it is clear that Equation (4.2) is continuous at $r = R_0$ since both sides of the function simply reduce to 0.5. However, the continuity of the derivatives of Equation (4.2) are not trivial. By considering the partial derivatives of Equation (4.2) with respect to each of the three fit parameters, the correlation between A and ΔR can be understood.

$$\frac{\partial LTP(r)}{\partial \Delta R} = \begin{cases} \frac{(r-R_0)e^{\frac{r-R_0}{\Delta R}}}{\Delta R^2\left(1+e^{\frac{r-R_0}{\Delta R}}\right)^2} & \text{for } r \leq R_0 \\ 0 & \text{for } r > R_0 \end{cases} \tag{4.3}$$

$$\frac{\partial LTP(r)}{\partial A} = \begin{cases} 0 & \text{for } r \leq R_0 \\ \frac{1}{2}\left(r - R_0\right) e^{A(r-R_0)} & \text{for } r > R_0 \end{cases} \tag{4.4}$$

$$\frac{\partial LTP(r)}{\partial R_0} = \begin{cases} \frac{e^{\frac{r-R_0}{\Delta R}}}{\Delta R\left(1+e^{\frac{r-R_0}{\Delta R}}\right)^2} & \text{for } r \leq R_0 \\ -\frac{1}{2}Ae^{A(r-R_0)} & \text{for } r > R_0 \end{cases} \tag{4.5}$$

Notice that Equations (4.3) and (4.4) both reduce to 0 at $r = R_0$ and are therefore continuous. However, imposing continuity at $r = R_0$ in Equation (4.5) reveals the condition:

$$A = -\frac{1}{2\Delta R} \tag{4.6}$$

This suggests a new fit function to replace Equation (4.2) and reduce the number of free fit parameters from three to two. The explicit form of the new fit function is given by

$$LTP(r) = \begin{cases} \frac{1}{1+e^{\frac{r-R_0}{\Delta R}}} & \text{for } r \leq R_0 \\ \frac{1}{2e^{\frac{r-R_0}{2\Delta R}}} & \text{for } r > R_0 \end{cases} \tag{4.7}$$

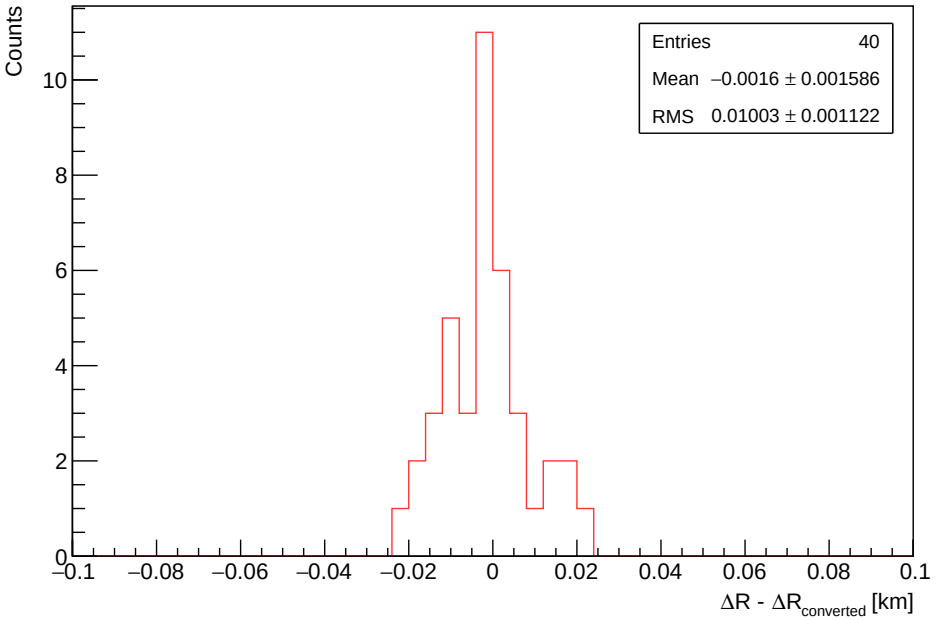


Figure 4.1. Distribution of the difference between the fitted ΔR and A converted to $\Delta R_{\text{converted}}$ using Equation (4.6). Note the mean difference is 1.6 m and the most extreme differences are approximately 20 m. This not only shows a strong correlation between the two fit parameters, but also shows that the fit function naturally fits ΔR and A to the same value.

where R_0 and ΔR are free fit parameters with R_0 being the distance where the LTP is equal to 0.5 and ΔR controlling the steepness of the LTP.

In order to investigate this relationship and correlation between ΔR and A in data, the full analysis procedure was completed with the original fit function, Equation (4.2) and the fit function with only two parameters, Equation (4.7). By converting the fitted values of A to an "equivalent" ΔR (denoted $\Delta R_{\text{converted}}$) using Equation (4.6), their values can be directly compared. A distribution of the difference, $\Delta R - \Delta R_{\text{converted}}$ is shown in Figure 4.1.

The differences are on the order of only meters, so the data clearly shows a strong correlation between ΔR and A. Given these results, the following analysis will use the new fit function, Equation (4.7) rather than the original fit function.

4.2.3 Parametrisation Results

The LTP for each radius bin up to 3 km in station-axis distance is calculated for various energy and zenith angle bins. The data are then fitted with Equation (4.7) in order to determine the fit parameters' dependence on energy and zenith angle.

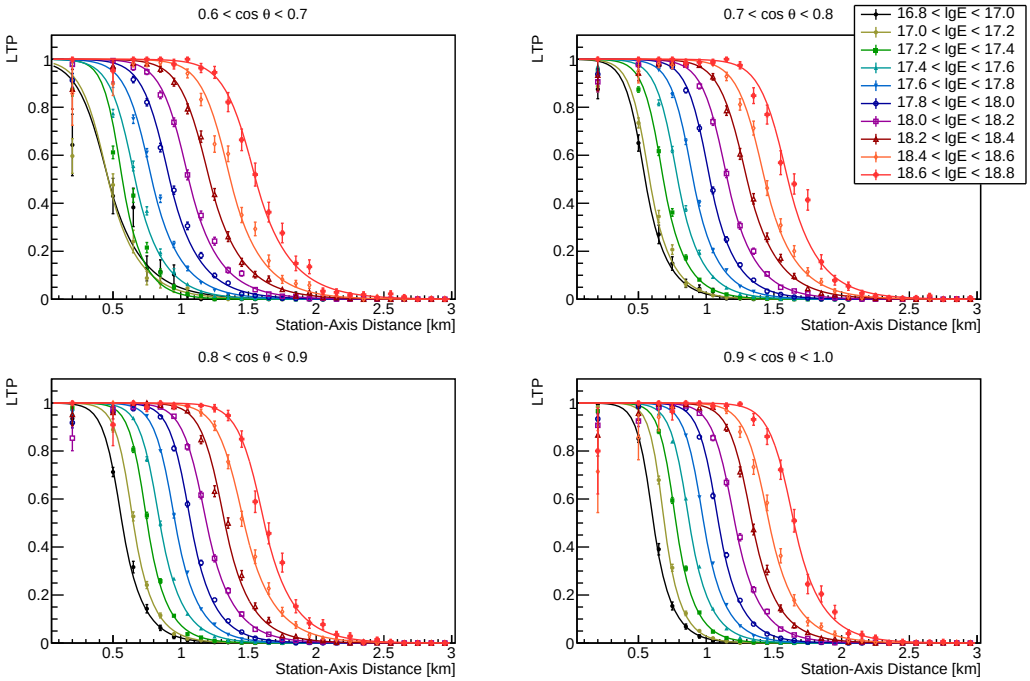


Figure 4.2. The measured LTP data and fits for ToT triggered stations.

This has been done for each of the trigger configurations listed in Section 4.2.1. An example of the LTP fit for ToT triggers only is shown in Figure 4.2.

The results of the LTP fits for the other trigger combinations are shown in Appendix A.2. As expected, the maximum effective distance for the trigger of a single detector increases with energy and, for a given energy with the cosine of zenith angle (i.e. vertical showers tend to trigger more often than showers with larger zenith angle). The dependence of the fit parameters on energy and zenith angle is investigated further in Section 4.2.4 below. Using these results, a final parametrisation of the average LTP as a function of energy and zenith angle is derived for each of the trigger configurations. A comparison of LTP parametrisations between the old and new triggers is shown in Section 4.2.6. Section 4.2.7 shows a comparison of the average LTP derived in this work with the LTP derived from simulations of pure proton and iron showers in [190]. The coefficients of the parametrisations for the average LTP are tabulated in Appendix A.1. The performance of the parametrisations is verified by an end-to-end comparison between the derived parametrisations and the data. The agreement is extremely good for most energy and zenith angle bins. These results are shown in Appendix A.4.

4.2.4 Parametrisation of Fit Parameters in Energy and Zenith Angle

The energy and zenith angle dependence of the fit parameters R_0 and ΔR have been studied in detail. In order to determine the combined dependence of energy and zenith angle on each of the fit parameters, a quadratic surface is fitted to each of the parameter values as functions of $\cos\theta$ and $\log E$. This approach reduces the uncertainties associated with fitting in energy and zenith angle separately. The following matrix equation shows the form of the fit for each of the parameters:

$$\text{Param} = \begin{pmatrix} 1 \\ \cos\theta \\ \cos^2\theta \end{pmatrix}^T \cdot \left[\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \log E \\ \log^2 E \end{pmatrix} \right]$$

where a to i are the free fit parameters. The results of these fits are shown in Appendix A.3 and the corresponding coefficients are tabulated in Appendix A.1 for each of the investigated trigger conditions. Each of the trigger configurations fit the quadratic surfaces very well and the fits produce a parametrisation which follows the data closely. See Appendix A.4 for a comparison between the data with each of the derived parametrisations.

Finally, each of the parameters R_0 and ΔR are plotted as a function of $\cos\theta$ for various energies in Figure 4.3. It is clear that ΔR varies by approximately a factor of 10 less than R_0 over the zenith angle and energy range considered in this study, and therefore has less of an impact on the final LTP. Thus, R_0 is the dominant parameter. As such, it provides an indication of the LTP without the need to evaluate the entire function. This will be discussed in more detail in Section 4.3.

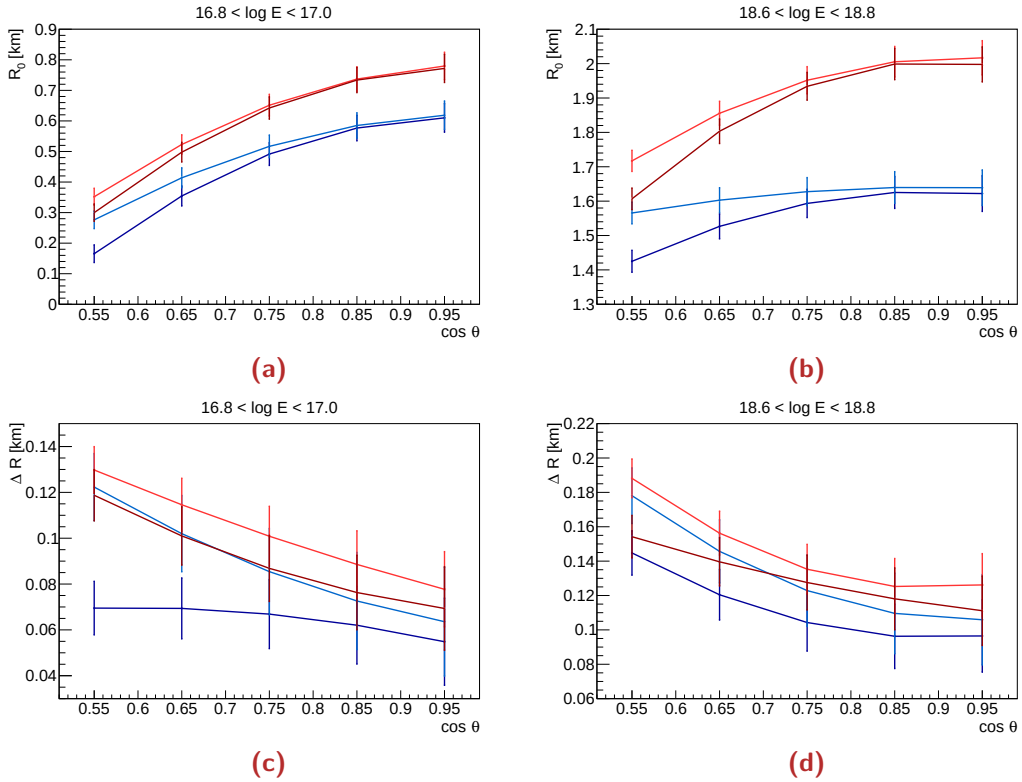


Figure 4.3. Zenith angle dependence of the LTP fit parameters, R_0 and ΔR for the new triggers (red) and old triggers (blue). The darker colours show the parameters for just ToT triggers and the lighter colours include TH triggers as well. (a) and (c) show the parameters for the range $16.8 < \log E < 17.0$. (b) and (d) show the parameters for the range $18.6 < \log E < 18.8$. R_0 strictly increases with the cosine of the zenith angle, that is, vertical showers always have a larger single station trigger probability than that of more inclined showers.

4.2.5 Error Propagation

As well as the expected LTP value, uncertainties are propagated from the initial fits using Equation (4.7) to the final parametrisations. The two fit parameters are assumed to be uncorrelated based on the study in Section 4.2.2, however correlations between fit parameters in subsequent steps are taken into account.

In general, the uncertainty of any function $f(a_i)$ (where a_i are free parameters) is given by the following equation (summing over the indices)

$$\sigma_f^2 = \frac{\partial f}{\partial a_i}\sigma_{a_i}\rho_{ij}\frac{\partial f}{\partial a_j}\sigma_{a_j} \tag{4.8}$$

where σ_{a_i} is the uncorrelated uncertainty of parameter a_i and ρ_{ij} is the correlation coefficient between parameter a_i and a_j.

In the parametrisation of the parameters R_0 and ΔR from Equation (4.7) discussed in Section 4.2.4, there are nine parameters fit in the process (a to i). Equation (4.8) is used to calculate uncertainties in R_0 and ΔR. Using R_0 as an example, which is of the form

$$R_0 = \begin{pmatrix} 1 \\ \cos\theta \\ \cos^2\theta \end{pmatrix}^T \cdot \left[\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \log E \\ \log^2 E \end{pmatrix} \right]$$

and substituting into Equation (4.8) gives

$$\sigma_{R_0}^2 = \frac{\partial R_0}{\partial a_i}\sigma_{a_i}\rho_{ij}\frac{\partial R_0}{\partial a_j}\sigma_{a_j} \tag{4.9}$$

where a_i are parameters a to i.

Then considering Equations (4.7) and (4.8), the uncertainty in the LTP is given by

$$\sigma_{LTP}^2 = \left(\frac{\partial LTP}{\partial R_0}\sigma_{R_0}\right)^2 + \frac{\partial LTP}{\partial R_0}\sigma_{R_0}\rho\frac{\partial LTP}{\partial \Delta R}\sigma_{\Delta R} + \left(\frac{\partial LTP}{\partial \Delta R}\sigma_{\Delta R}\right)^2 \tag{4.10}$$

where the correlation ρ, between R_0 and ΔR is assumed to be negligible. This gives a final uncertainty in the LTP of:

$$\sigma_{LTP}^2 = \left(\frac{\partial LTP}{\partial R_0}\sigma_{R_0}\right)^2 + \left(\frac{\partial LTP}{\partial \Delta R}\sigma_{\Delta R}\right)^2 \tag{4.11}$$

4.2.6 Comparison Between Old and New Triggers

Each of the derived LTP parametrisations can be compared in order to study the trigger probability for different trigger conditions (outlined in Section 3.1.2). In this case, it is most interesting to study the relative increase in the trigger probability when including the new triggers (ToTd and MoPS) compared to only the old triggers (ToT and Th). Figure 4.4 clearly shows a significant increase in the trigger probability when including the new triggers compared with the old triggers. This is to be expected since the new triggers are designed to trigger on low signals produced by multiple particles spread out in time rather than a large signal produced by a single muon, for example. These two cases are distinguished by the modified algorithms which are implemented by the two new triggers.

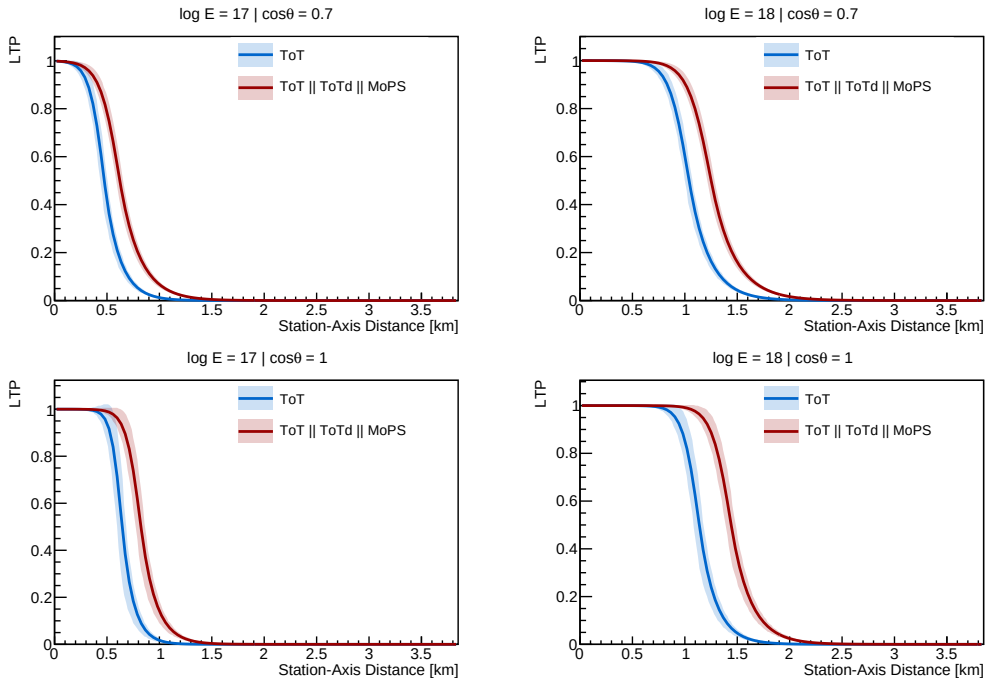


Figure 4.4. An example comparison for the LTP derived from hybrid data with only ToT triggers (blue) with ToT || ToTd || MoPS triggers (red). Note the clear increase in the trigger probability when the new triggers are incorporated. The shaded regions represent the 1σ confidence region based on propagation of fit uncertainties (described in Section 4.2.5).

4.2.7 Comparison with Simulation

Multiple studies using simulated CORSIKA showers have been made to derive LTP parametrisations. It is appropriate to compare the derived parametrisations from this work to the parametrisation derived from simulations in order to validate both results. Figure 4.5 shows a comparison between the ToT parametrisation derived from hybrid data in this work and proton and iron parametrisations derived from simulations in [190]. The agreement between the parametrisations depends on the energy and zenith angle, however any differences are relatively small. It is recommended to use the data derived LTP parametrisation for calculating fiducial cuts in X_{max} analysis for example, since the data naturally contains the true mass composition information and avoids the need to extrapolate hadronic models.

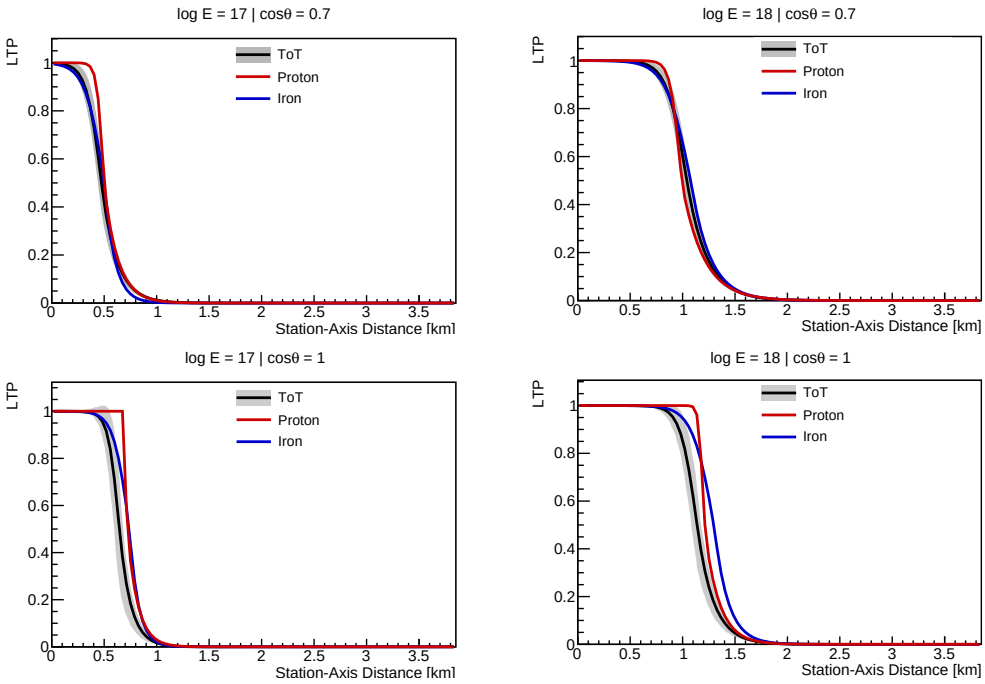


Figure 4.5. An example comparison of the LTP derived using hybrid data with only ToT triggers (black) to simulations from [190] with proton (red) and iron (blue) parametrisations shown. The shaded region represents the 1σ confidence region based on propagation of fit uncertainties (described in Section 4.2.5). Note the close agreement between the data and simulations.

4.3 Array Trigger Efficiency from Hybrid Data

In order to investigate the array trigger probability as a function of energy, a toy MC simulation of an ideal surface detector array is used along with the data derived LTP parametrisation from this study to determine the trigger efficiency of both the old and new SD triggers. This is done for both single station triggers (the requirement for hybrid events) and for SD T4 triggers (requiring at least 3 ToT triggers). In order to evaluate the trigger condition in the SD case, the 3ToT and 4C1 algorithms have been re-implemented from the Offline module, `SdEventSelector`.

4.3.1 Analysis Method

In this analysis, 100,000 events have been randomly simulated inside of a triangle of stations. This has been done in discrete steps of energy from $10^{16} - 10^{19}$ eV and for various zenith angles up to 60° for both the 750 m array and the 1500 m array. For each event, the geometry of the shower is sampled randomly in three stages: Firstly, a random core position inside a central triangle of stations is selected. An example of the core positions of 1000 events in the 1500 m array are shown in Figure 4.6. Then an azimuth, ϕ between 0° and 360° is selected from a uniform distribution. Finally, the zenith angle, θ is then sampled from the distribution $\sin\theta\cos\theta$.

Next, the station-axis distance between the generated shower axis and each of the stations in the array shown in Figure 4.6 is calculated, and the LTP is evaluated at the given energy, station-axis distance and zenith angle. A trigger is then decided by a "hit-or-miss" method. That is, generating a random number between 0 and 1, if the evaluated trigger probability is larger than the generated random number then the station is considered to have triggered. The array trigger efficiency at a given energy is then defined as the ratio of the number of triggered events to the total number of events. The trigger efficiency is also evaluated using the $\pm 1\sigma$ LTP as discussed above in Section 4.2.5, in order to propagate 1σ uncertainties from the LTP parametrisations through to the trigger efficiency and energy thresholds for full efficiency.

4.3.2 Comparison of Old and New Triggers

This section compares the trigger efficiency of vertical events ($\theta < 60^{\circ}$) with and without the new triggers with the expectation of observing a reduced energy threshold when the new triggers are included. Figures 4.7 and 4.8 show the trigger efficiency as a function of $\log E$ for single station triggers and for SD T4 triggers (a 3ToT or 4C1 SD event trigger is required), respectively.

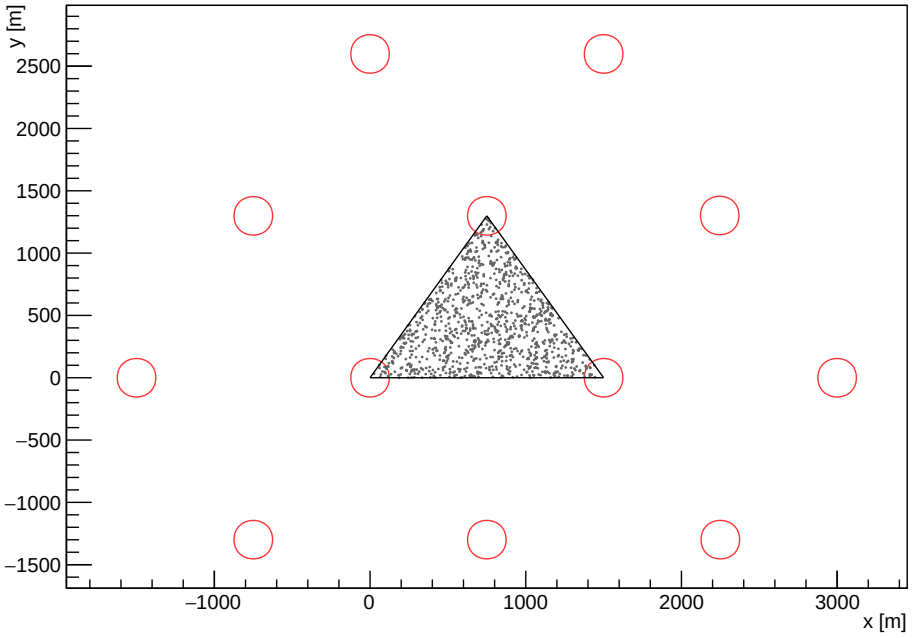


Figure 4.6. An example of the random core positions sampled inside of an ideal 1500 m array. This layout of stations is used throughout this study. All core positions are sampled inside the central triangle of stations in order to avoid edge effects introduced by shower cores located outside the array or close to the edge, where the trigger probability is reduced.

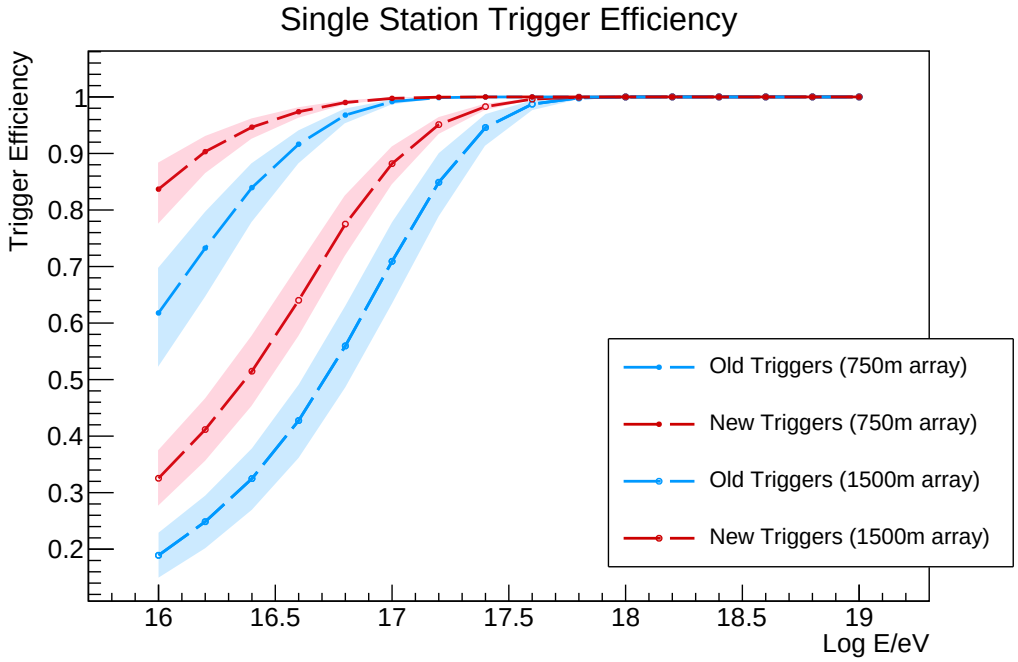


Figure 4.7. The trigger efficiency as a function of $\log E$ over all zenith angles up to $60°$ for single station triggers. It is clear the new triggers (red) show a significant reduction in the energy threshold for full efficiency compared to the old triggers (blue). The shaded regions represent the 1σ confidence region based on propagation of LTP fit uncertainties (described in Section 4.2.5).

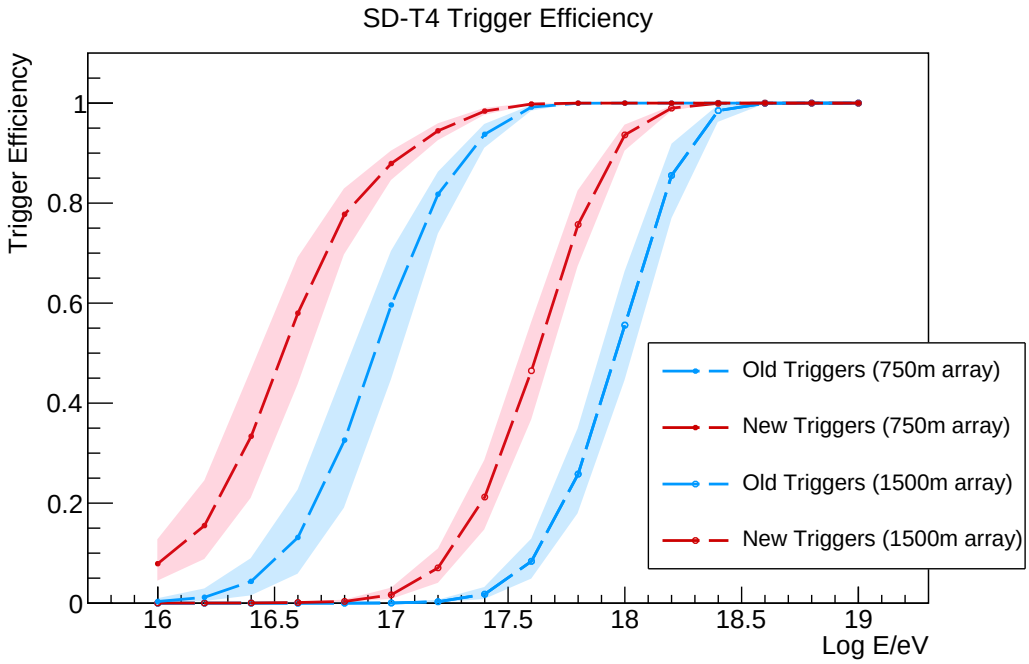


Figure 4.8. The trigger efficiency as a function of $\log E$ over all zenith angles up to 60° for SD triggers. It is clear the new triggers (red) show a significant reduction in the energy threshold for full efficiency compared to the old triggers (blue). The shaded regions represent the 1σ confidence region based on propagation of LTP fit uncertainties (described in Section 4.2.5).

In the case of the new triggers (ToTd and MoPS) a 3ToT station can be any of 3ToT || 3ToTd || 3MoPS. The results show a significant reduction in energy threshold when including the new triggers compared to only the old triggers in both the single station and the SD T4 case. It is important to note that the LTP parametrisations only include data down to $10^{16.8}$ eV and therefore any interpretation of the trigger efficiency below this energy is an extrapolation of the parametrisation and cannot be considered reliable. However, the key finding of this study is the determination of the energy threshold for full trigger efficiency, which is larger than or very close to $10^{16.8}$ eV in general. This requires very little, or no extrapolation of the LTP parametrisations, so the energy thresholds can therefore be considered reliable.

For hybrid events only requiring a single SD station to trigger, the detector is fully efficient (defined as 98% efficiency) at $10^{16.7}$ eV for the new triggers and $10^{16.9}$ eV for the old triggers in the 750 m array. For the 1500 m array, the detector is fully efficient at $10^{17.4}$ eV and $10^{17.55}$ eV for the new and old triggers, respectively. This is a significant reduction in the energy thresholds. Similarly, for SD-T4 events (3ToT || 4C1), the detector is fully efficient at $10^{17.4}$ eV for the new triggers and $10^{17.5}$ eV for the old triggers in the 750 m array, and $10^{18.15}$ eV for the new triggers and $10^{18.4}$ eV for the old triggers in the 1500 m array. These energy threshold values are summarised in Tables 4.1 and 4.2.

Trigger	$\theta < 40^\circ$	$\theta < 55^\circ$	$\theta < 60^\circ$
Hybrid (new)	16.2	16.5	16.7
Hybrid (old)	16.5	16.7	16.9
SD-T4 (new)	16.85	17.15	17.4
SD-T4 (old)	17.15	17.35	17.5

Table 4.1. Energy Threshold [$\log_{10} E$/eV] for the 750 m array with the three integral zenith angle ranges.

Trigger	$\theta < 40^\circ$	$\theta < 55^\circ$	$\theta < 60^\circ$
Hybrid (new)	17.1	17.25	17.4
Hybrid (old)	17.4	17.5	17.55
SD-T4 (new)	17.95	18.05	18.15
SD-T4 (old)	18.38	18.385	18.4

Table 4.2. Energy Threshold [$\log_{10} E$/eV] for the 1500 m array with the three integral zenith angle ranges.

4.3.3 Zenith Angle Dependence

In order to study the effect that zenith angle cuts have on the energy threshold for full efficiency, the efficiency is plotted as a function of $\log E$ for three different integral zenith ranges ($< 40^\circ$, 55° and 60°) for both the old and new triggers in Figure 4.9. There is a significant zenith angle dependence in the trigger efficiency in all cases for the 750 m array. However, the efficiency using the old triggers (Figures 4.9a and 4.9c) does not appear to have any significant zenith angle dependence for the 1500 m array. All energy thresholds for full efficiency are summarised in Tables 4.1 and 4.2.

In order to study the zenith angle dependence of the trigger efficiency directly, the efficiency is plotted as a function of $\log E$ for several differential zenith angles from $0-60^\circ$. This has been done for both the old and new triggers in Figures 4.10 and 4.11. It seems that in general vertical showers have a higher trigger efficiency. That is, for a fixed energy, vertical showers are more likely to trigger than inclined showers. This is especially clear for the single station trigger (hybrid events). However, the effect is less significant for the array trigger (3ToT || 4C1). In fact, for the old triggers in the 1500 m array, vertical events have a higher energy threshold for full efficiency than 40° inclined showers. To elucidate this effect, it is important to consider that there are two competing effects which need to be balanced. Firstly, vertical showers are expected to trigger at lower energies since the electromagnetic component of the shower is strongest compared to inclined showers, where these particles travel through more atmospheric depth and hence suffer more attenuation. However, there is a second effect where the distance between stations in the shower plane become

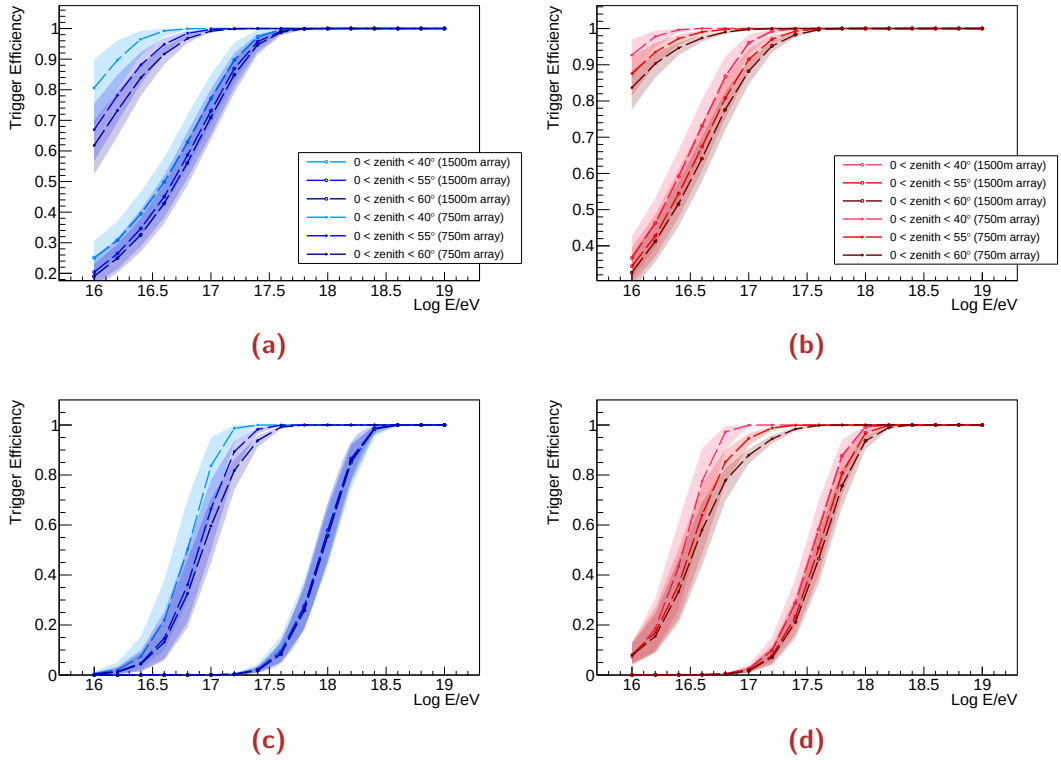


Figure 4.9. The effect of successive zenith angle cuts at $40°$, $55°$ and $60°$. Each plot shows the efficiency as a function of energy for each zenith angle cut for both the 750 m and 1500 m array. (a) and (b) show the single station efficiency for the new and old triggers, respectively. (c) and (d) show the SD T4 efficiency. The shaded regions represent the 1σ confidence region based on propagation of LTP fit uncertainties (described in Section 4.2.5).

effectively smaller for inclined showers. This means the trigger probability of any one station increases as if the stations were denser; hence a lower energy threshold. It is not clear how these two effects are expected to balance; however, it appears this transition from vertical domination of the energy threshold to inclined domination has been observed in Figure 4.11c. It should be made clear that this effect is solely due to the array trigger algorithm and not an upward bias in the LTP of inclined showers compared to vertical showers. This is clear in Figures 4.3a and 4.3b where the parameter R_0 is plotted as a function of zenith angle for various energies and in Figure 4.10 for single station triggers where the effect is not observed, as would be expected if this scenario is correct.

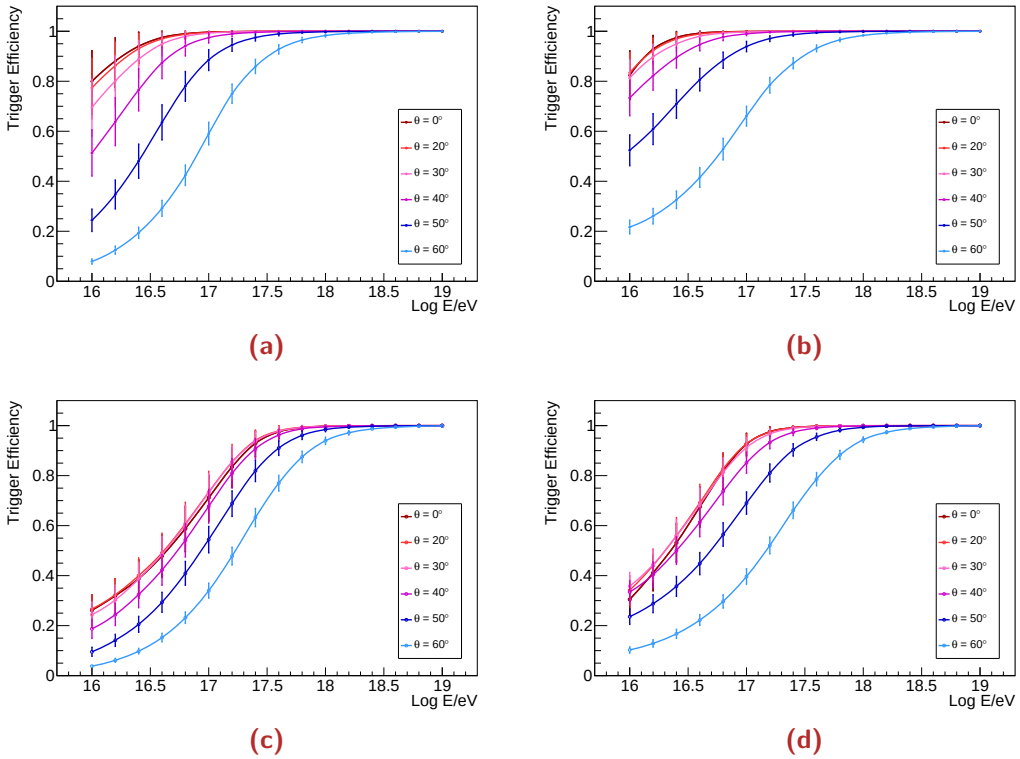


Figure 4.10. Zenith angle dependence of the single-station trigger efficiency from 0° to 60° zenith angle. Red shows the efficiency for vertical showers and blue shows progressively larger zenith angles. (a) and (b) show the efficiency for the old and new triggers, respectively, for the 750 m array. (c) and (d) show the efficiency for the 1500 m array. The error bars represent the 1σ confidence region based on propagation of LTP fit uncertainties (described in Section 4.2.5).

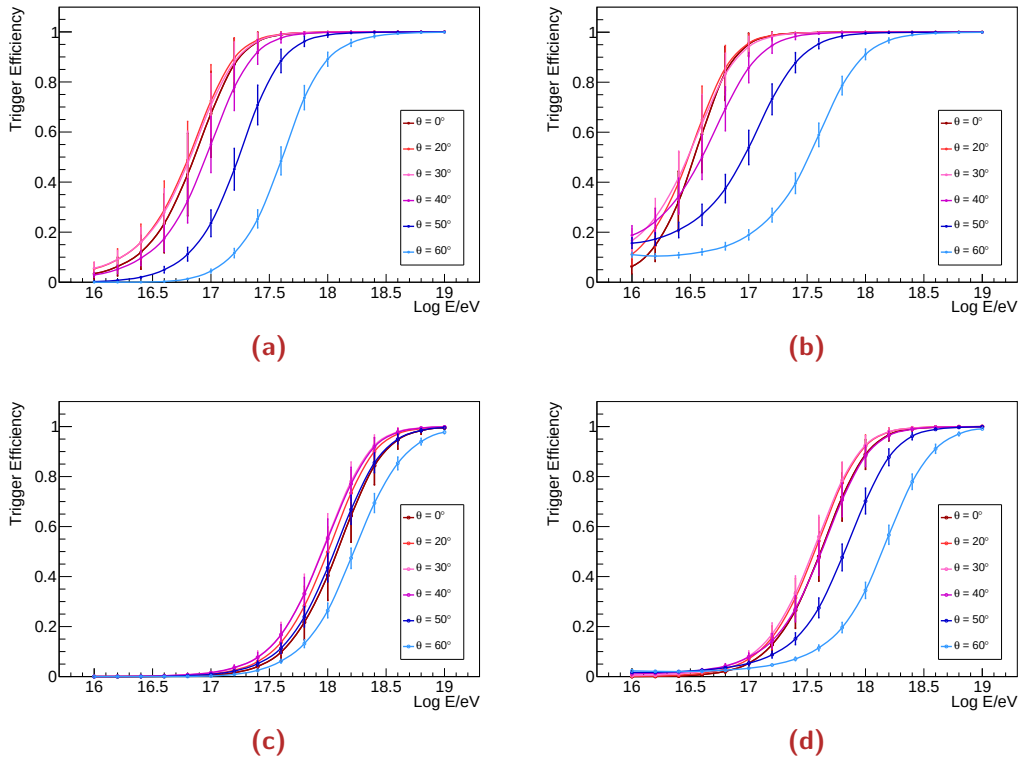


Figure 4.11. Zenith angle dependence of the SD T4 trigger efficiency from 0° to 60° zenith angle. Red shows the efficiency for vertical showers and blue shows progressively larger zenith angles. (a) and (b) show the efficiency for the old and new triggers, respectively, for the 750 m array. (c) and (d) show the efficiency for the 1500 m array. The error bars represent the 1σ confidence region based on propagation of LTP fit uncertainties (described in Section 4.2.5).

4.3.4 Comparison with Simulation Studies

Several studies of the SD array trigger efficiency using simulations have been conducted in the past using a variety of methods. This section compares 3ToT trigger efficiencies for the new triggers (i.e. 3ToT || 3ToTd || 3MoPS) from simulation studies at three different integral zenith angle ranges up to 40°, 55° and 60° for the 750 m array in Figures 4.12 to 4.14. The results from [141, 170] show the 3ToT efficiency for the new triggers. The $\pm 1\sigma$ error bars from this work encapsulates all of these simulations in the relevant energy range (above $10^{16.8}$ eV), with the exception of the data from [191]. The large difference may be due to the choice of hadronic model (QGSJetII-03 in this case) or an energy scale difference. There is only a small difference between hadronic models and composition types for the other two simulation studies; they use QGSJetII-04 and EPOS hadronic models [141, 170]. For vertical events in Figure 4.12 there is strong agreement between the efficiency from data and simulations above $10^{16.8}$ eV, below which the LTP has been extrapolated to lower energies.

There is a clear bias in the data efficiency in the extrapolation region, however since this is well below the energy threshold for full efficiency it has no detrimental effect. Additionally, due to this bias, the data compares better with simulations of lighter composition at larger zenith angles. For the energy range above 10^{17} eV, the data is consistent with all simulations. This bias could exist for a variety of reasons such as an excess of muons in the data which is not accounted for in simulations. The bias may also be due to the simplistic treatment of the array in this study; the array described in Section 4.3.1 is effectively an ideal, infinite array with no edges. Simulations may treat the array edges differently. This was not considered further as the difference between data and simulation is small, especially since the agreement is still very good in the region above 90% efficiency, which is of greatest interest for this study.

Since the agreement with simulations is good, it gives more confidence in both the simulations and the trigger efficiency derived in this work. Additionally, the agreement suggests that the procedure used in this work can be used as a new reliable method for estimating the trigger efficiency of any configuration of the SD without the need for computationally expensive simulations. Since the LTP has already been derived in this work, the trigger efficiency of any configuration of SD stations can be estimated very quickly with the toy MC simulation used in this study of an ideal array. In principle, this work could be extended to estimate the efficiency of the true surface array using the same procedure.

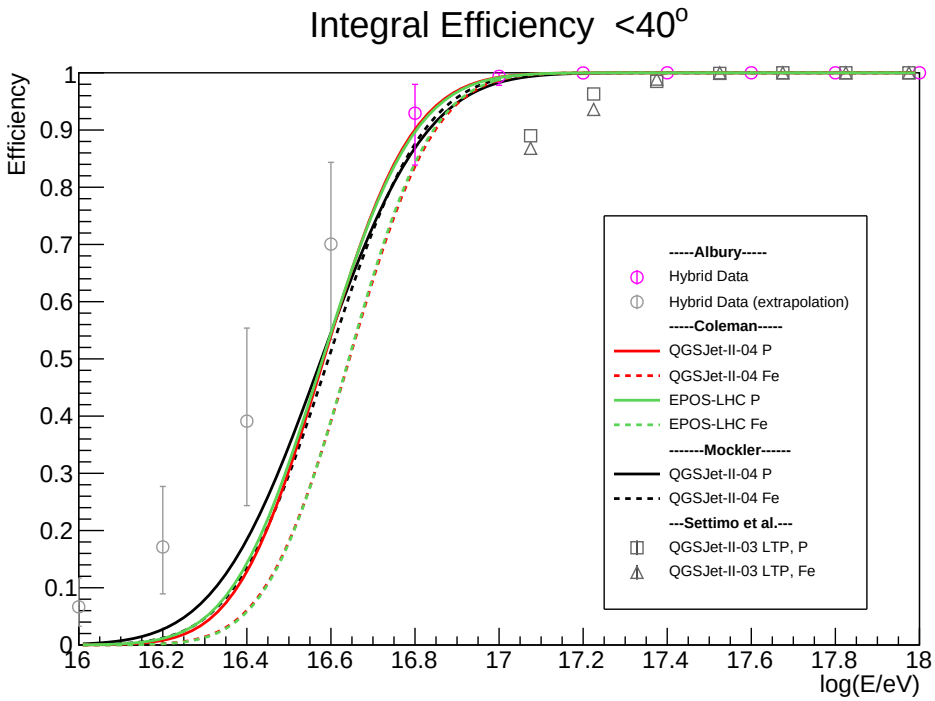


Figure 4.12. Comparison of LTP derived efficiency (this work) with multiple simulation studies up to zenith angles of $40°$ for the 750 m array. Good agreement is observed, especially in the region above an efficiency of 0.9. The error bars represent the 1σ confidence region based on propagation of LTP fit uncertainties (described in Section 4.2.5).

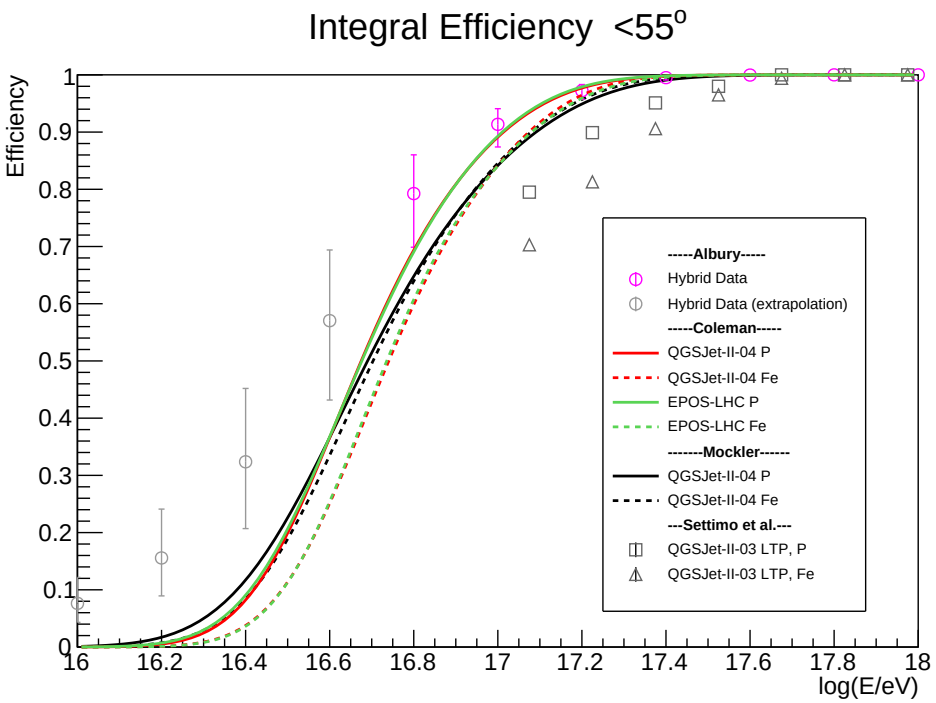


Figure 4.13. Comparison of LTP derived efficiency (this work) with multiple simulation studies up to zenith angles of $55°$ for the 750 m array. Good agreement is observed near full efficiency, however a small systematic deviation between data and the bulk of simulations is present. The error bars represent the 1σ confidence region based on propagation of LTP fit uncertainties (described in Section 4.2.5).

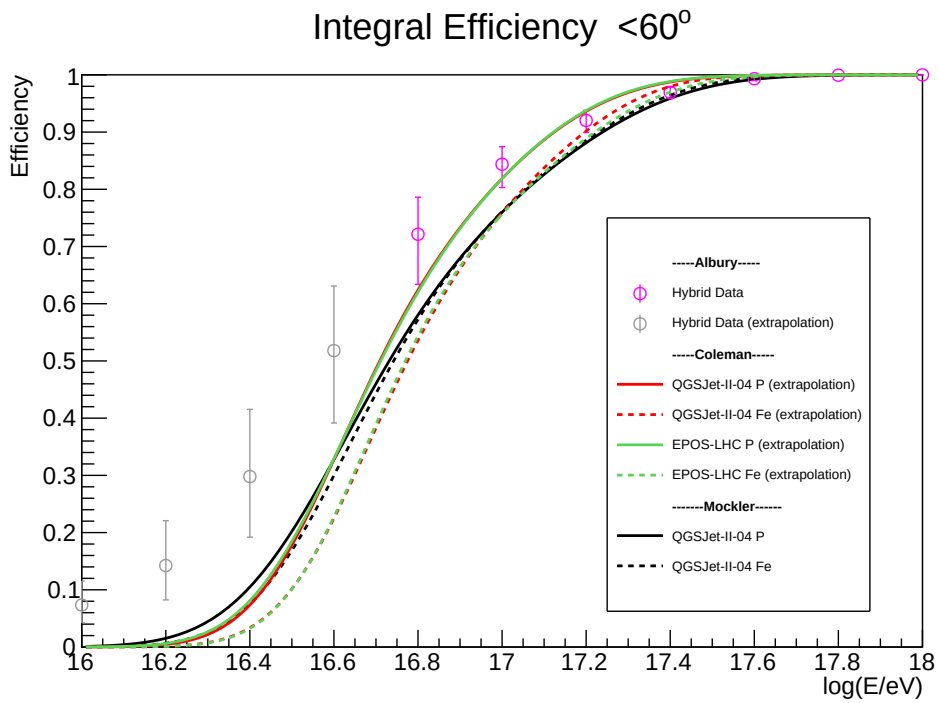


Figure 4.14. Comparison of LTP derived efficiency (this work) with multiple simulation studies up to zenith angles of $60°$ for the 750 m array. Good agreement is observed near full efficiency, however a small systematic deviation between data and the bulk of simulations is present. The error bars represent the 1σ confidence region based on propagation of LTP fit uncertainties (described in Section 4.2.5).

4.4 Conclusions

The LTP was derived using hybrid data and parametrised as a function of zenith angle and energy from $10^{16.8}$ to 10^{19} eV. Good agreement was found between the derived parametrisations in this work and from previous simulation studies. It is recommended that the data derived parametrisations are used for fiducial selection cuts in the X_{max} analysis. The data derived parametrisations naturally contain the true mass composition information, and do not rely on details of hadronic interaction models.

The array trigger efficiency for both single station and array triggers were studied in detail using a toy MC simulation. Using an ideal SD array and the LTP derived from data in this work, the threshold for full trigger efficiency was investigated. The effect of the new SD triggers is evident; it extends the full trigger efficiency to lower energies. A comparison with the trigger efficiency in this work and from several simulation studies showed good agreement, especially in the region above $10^{16.8}$ eV. The zenith angle dependence of the trigger efficiency was also investigated. Different zenith angle cuts were evaluated to improve the efficiency (e.g. only include more vertical showers) as well as to study the effect of the triggers as a function of zenith angle. The largest improvement between the old and new triggers is for near-vertical showers. There is also evidence to suggest there may be a "sweet-spot" for the trigger efficiency, where the most efficient zenith angle is slightly inclined rather than exactly vertical. This effect has not been observed in simulation studies.

The results of this study confirm a reduced energy threshold with the new triggers. This suggests that the HeCo X_{max} analysis could, in principle, extend down to $10^{16.7}$ eV without a composition bias in the SD trigger. This energy threshold can be further reduced (if required) with appropriate zenith angle cuts. In order to benefit from a lower energy SD trigger threshold, the X_{max} analysis (at energies lower than $\sim 10^{17.5}$ eV) will need to consider utilising detector simulations to account for the FD X_{max} acceptance. An extension of the X_{max} analysis to lower energy using the new SD triggers is investigated in Chapter 5.

5 Mass Composition Studies

5.1 Introduction

As discussed in Section 2.5, the mass composition of ultra-high energy cosmic rays (UHECR) is an important observable which can be used to understand features of the energy spectrum and arrival directions of UHECRs. The data collected by the Fluorescence Detectors (FDs) of the Pierre Auger Observatory (Auger) are used to measure the depth of shower maximum, X_{max}. While event-by-event discrimination of mass composition is not possible at the highest energies due to large shower-to-shower fluctuations, the distribution of observed X_{max} can provide estimates of the average mass composition using the predictions of various hadronic models. In order to achieve an accurate estimate of the average mass composition from X_{max}, high quality, unbiased data is essential. In the standard X_{max} analysis used by Auger, strict data cuts are used to ensure these conditions. These cuts result in a significant reduction in the number of events used in the analysis. This work is focused on the development of a new technique which can ensure unbiased data, while maintaining a large data sample. The main motivation for increasing the number of events used in the analysis is to extend the energy range of the analysis which currently has a lower energy threshold of $10^{17.2}$ eV. This threshold is limited by the trigger efficiency of the Surface Detector (SD) as well as the field of view (FoV) cuts which are described in Section 5.2.1. Following the work in Chapter 4, it will be possible to extend the X_{max} analysis to lower energies using the new SD triggers described in Section 3.1.2.

The X_{max} analysis is separated into two subsets; the HEAT-Coihueco (HeCo) data set, and the standard FD data set. The HeCo data set, as its name suggests, includes data measured by the High-elevation Auger telescopes (HEAT) and Coihueco FD telescopes. These measurements take advantage of the low energy extensions of Auger including the 750 m array of SD stations and the higher elevation of HEAT telescopes. The standard FD data set includes data above $10^{17.6}$ eV from the four standard FDs; Los Leones, Los Morados, Loma Amarilla and Coihueco. The energy threshold of the standard FD data set is limited by the trigger efficiency of the SD in the standard 1500 m configuration.

The work in this chapter builds upon the original ideas from Plum [192]. A modified X_{max} analysis technique is developed in this work. Detailed studies of the reconstruction bias and detector resolution are used in conjunction with Monte Carlo (MC) simulations to thoroughly test the newly developed analysis technique, and systematic uncertainties and bias corrections are discussed in detail. The final X_{max} moments determined with this analysis are compared with the predictions of multiple hadronic models, as well as to the moments determined with the standard X_{max} analysis and those from other experiments.

5.2 Standard X_{max} Analysis

The modified X_{max} analysis technique developed in this work is best described in contrast to the standard analysis technique. In this section, the standard X_{max} analysis technique is described. In order to produce an unbiased data sample, the detector X_{max} acceptance must be carefully taken into account. The FoV cuts and subsequent acceptance corrections aim to produce an unbiased X_{max} distribution which can be compared directly to the predictions of various hadronic interaction models for nuclei of different masses.

5.2.1 Field of View Selection Cuts

In any experiment, an observable will only be able to be measured if it falls within the acceptance range of the detector. In the case of X_{max} measurements, the standard FD telescopes are geometrically limited by a physical FoV between elevation angles of 1.5° and 31.5°. Only X_{max} values measured within in the FoV of the telescopes will be directly observed. This FoV effect introduces a bias in the collected data sample used to measure X_{max} since the X_{max} distributions can become truncated. In order to remove this bias, selection cuts must be applied to the data to ensure only good quality, unbiased events are used in analysis. There are two main selection cuts applied to the data in addition to standard quality cuts. The first cut requires that the observed X_{max} is inside the "Expected" FoV (Section 5.2.1.1). The second FoV cut ensures showers with deep or shallow X_{max} are detected with equal probability, with deep showers having an X_{max} greater than $\sim$1000 g/cm^2, and shallow showers having an X_{max} $\sim$400 g/cm^2 or less. Due to the orientation of the FD telescopes, the probability of detecting showers with deeper or shallower X_{max} is not equal. Since different nuclei have different X_{max} distributions, this can result in a composition bias (see Section 5.2.1.2).

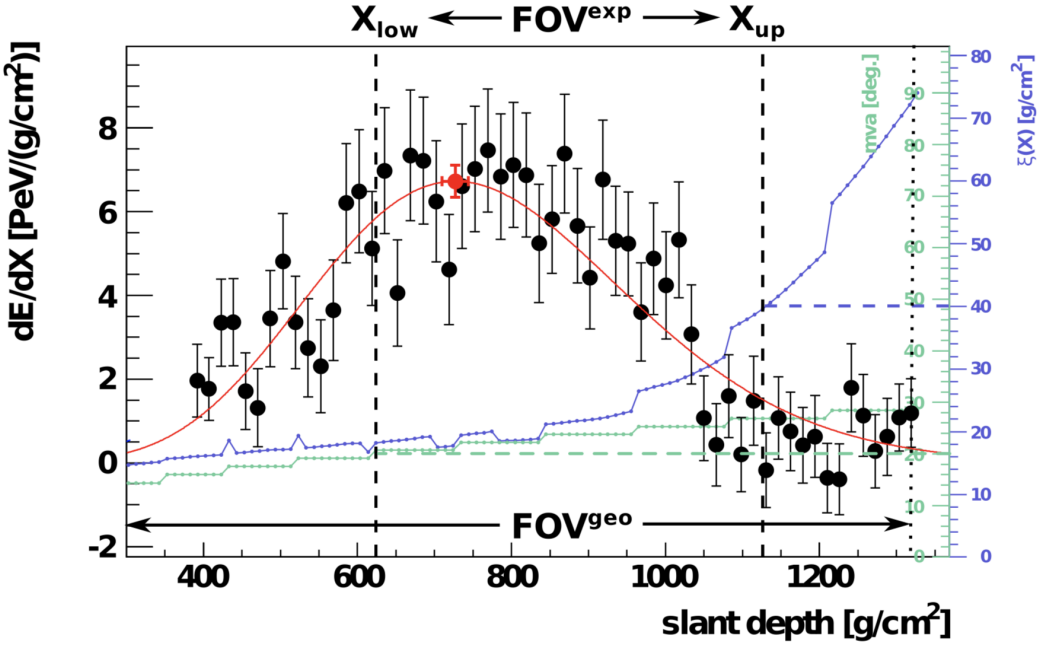


Figure 5.1. Expected $X_{\max}$ uncertainty (blue) derived from comparison with artificial Gaisser-Hillas shower profiles by using error propagation from geometry and profile reconstruction for an example event from Los Morados ($\log(E/\text{eV}) = 18.2$). The red line shows the fitted Gaisser-Hillas profile to the measured data (black points) with $X_{\max}$ indicated by the red point. Also shown is the minimum viewing angle in green as a function of slant depth. The expected FoV is defined by the range of depth where both the expected $X_{\max}$ uncertainty is less than 40 g/cm^2, and the minimum viewing angle is less than 20° [193].

5.2.1.1 Expected field of view

The physical *geometrical field of view* is simply defined by the intersection of the shower-detector plane (SDP) with the borders of the telescope cameras. More useful, however, is the *expected field of view* which is defined to ensure $X_{\max}$ is reconstructed with acceptable resolution. For example, a shower may be within the geometrical FoV of a telescope, but too far away to be measured by the FD telescopes. The expected FoV is defined independently for each event using MC simulated events. The observed shower profile is replaced with a Gaisser-Hillas profile, keeping the reconstructed energy and geometry fixed. The position of $X_{\max}$ is artificially varied in discrete steps to estimate the expected $X_{\max}$ uncertainty as a function of $X_{\max}$. An example is shown in Figure 5.1. Once this process has been completed, a range of $X_{\max}$ values, X_{low} and X_{up} can be derived for which the expected uncertainty is acceptable. For low $X_{\max}$ uncertainty there is a good correlation between the expected and measured uncertainty on $X_{\max}$, however at approximately 40 g/cm^2 the expected $X_{\max}$ uncertainty becomes systematically biased compared with the measured uncertainty [97]. Therefore, only events with estimated $X_{\max}$ uncertainty less than 40 g/cm^2 pass the expected FoV cut. In addition to the expected $X_{\max}$ uncertainty, the *viewing angle* between each pixel pointing

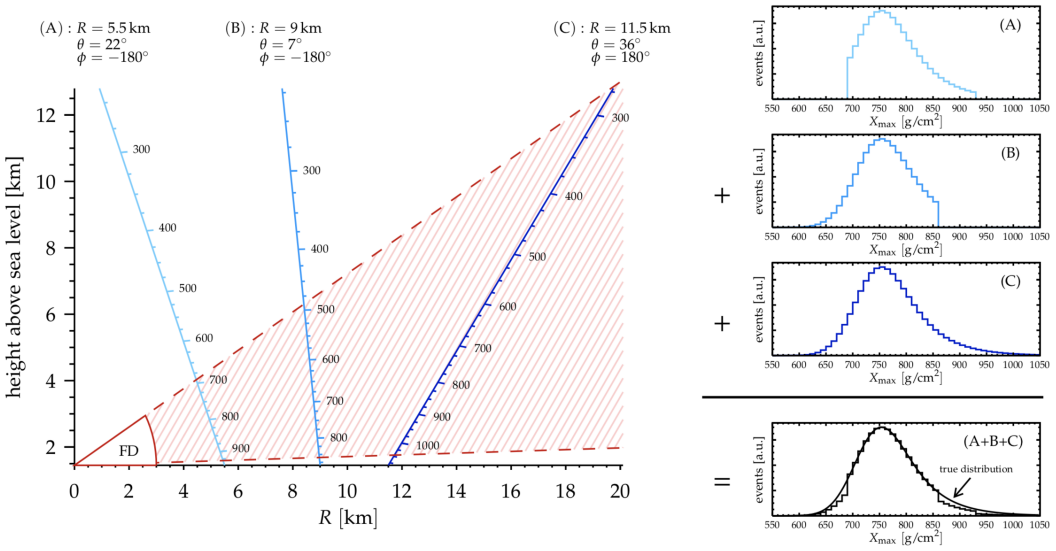


Figure 5.2. The effect of the FD FoV on the measurements of $X_{\max}$ and the resulting truncation of the $X_{\max}$ distribution depending on the shower geometry. Three contrasting cases are shown. In case A and B, the $X_{\max}$ distribution is truncated by the FoV. In case C, the entire $X_{\max}$ distribution is contained within the detector FoV. Examples of the corresponding $X_{\max}$ distributions and their sum is shown on the right. Note that the combined distribution is missing events from the tails of the $X_{\max}$ distribution, effectively providing the values close the peak with a larger weight. The FoV cuts and subsequent acceptance correction aim to avoid this effect [97].

direction and the shower axis provide another important quality measure. A small viewing angle corresponds to a large contribution from Cherenkov light. Such events are susceptible to systematic errors in the reconstruction of the shower geometry and thus deemed unreliable in the standard FD analysis. It is therefore necessary to require a minimum viewing angle of 20° in all pixels for a given event to be retained after selection cuts [97].

5.2.1.2 Fiducial field of view

The observed $X_{\max}$ distribution depends on the composition of the primary particles. On average, light nuclei penetrate the atmosphere deeper than heavy nuclei. Due to the limited FoV of the FD telescopes, a lighter primary particle with a deeper $X_{\max}$ may be more probable than an event with identical energy and geometry, but with a heavier primary particle. This means that only some subset of the true $X_{\max}$ distribution can be observed. This leads to a FoV composition bias where the $X_{\max}$ distribution becomes truncated. This effect is illustrated in Figure 5.2. Ideally the entire $X_{\max}$ distribution will be contained within the detector FoV (case C in Figure 5.2). The *fiducial field of view* cut is applied to select only events for which this truncation does not significantly change $\langle X_{\max} \rangle$ and ensures showers with deep or shallow $X_{\max}$ are detected with equal probability. Since the true $X_{\max}$

distribution is unknown, a data driven method is used to obtain the fiducial range where the distribution is unaffected. If the distribution is truncated (case A or B in Figure 5.2) then the mean of the X_{max} distribution will become systematically biased. The measured X_{max} distribution can be approximated by a convolution between a Gaussian and exponential distribution. In the perfect case, where no FoV limits exist, the mean of the X_{max} distribution is given by

$$\langle X_{\text{max}}^{\infty} \rangle = \frac{\int_0^{\infty} xG(x) \otimes E(x)dx}{\int_0^{\infty} G(x) \otimes E(x)dx}. \tag{5.1}$$

However, by introducing a FoV limit (integration limits), the distribution may become truncated, leading to the truncated mean

$$\langle X_{\text{max}}^{\text{trunc}} \rangle = \frac{\int_{X_{\text{low}}}^{X_{\text{up}}} xG(x) \otimes E(x)dx}{\int_{X_{\text{low}}}^{X_{\text{up}}} G(x) \otimes E(x)dx}. \tag{5.2}$$

In the case that X_{low} and X_{up} are far away from the edges of the X_{max} distribution, then the truncated mean is unbiased and therefore a good approximation of the true mean ($\langle X_{\text{max}}^{\text{trunc}} \rangle \approx \langle X_{\text{max}}^{\infty} \rangle$). Therefore, the mean of the X_{max} distribution as a function of the FoV limits, X_{low} and X_{up} can be used to determine the range for which the distribution is unaffected. An example of this is shown in Figure 5.3 for the energy range $18.1 < \log E < 18.2$. The fiducial FoV limits $X_{\text{low}}^{\text{fid}}$ and $X_{\text{up}}^{\text{fid}}$ are chosen in order to fulfil the condition

$$| \langle X_{\text{max}} \rangle - \langle X_{\text{max}}^{\infty} \rangle | < \Delta \tag{5.3}$$

where $\Delta = 5\text{g/cm}^2$ is chosen as the maximum allowed deviation [97]. The fiducial FoV limits are shown in Figure 5.3. The fiducial FoV limits are determined for each energy bin and energy dependent parametrisations of the fiducial limits are derived. The parametrisation has the form

$$f(E) = \begin{cases} a + b\,(\log E - E_0)^2 & \log E \leq E_0 \\ c & \text{otherwise} \end{cases} \tag{5.4}$$

for both the upper and lower fiducial limits. The parametrisation is shown along with the fiducial FoV limits for each energy bin as a function of energy in Figure 5.4. This parametrisation forms the final fiducial FoV cut. An event is then only retained if it passes the following conditions

$$X_{\text{low}} < X_{\text{low}}^{\text{fid}} \tag{5.5}$$

$$X_{\text{up}} > X_{\text{up}}^{\text{fid}} \tag{5.6}$$

Note that the cut is applied to the event-specific values of X_{low} and X_{up}. This means

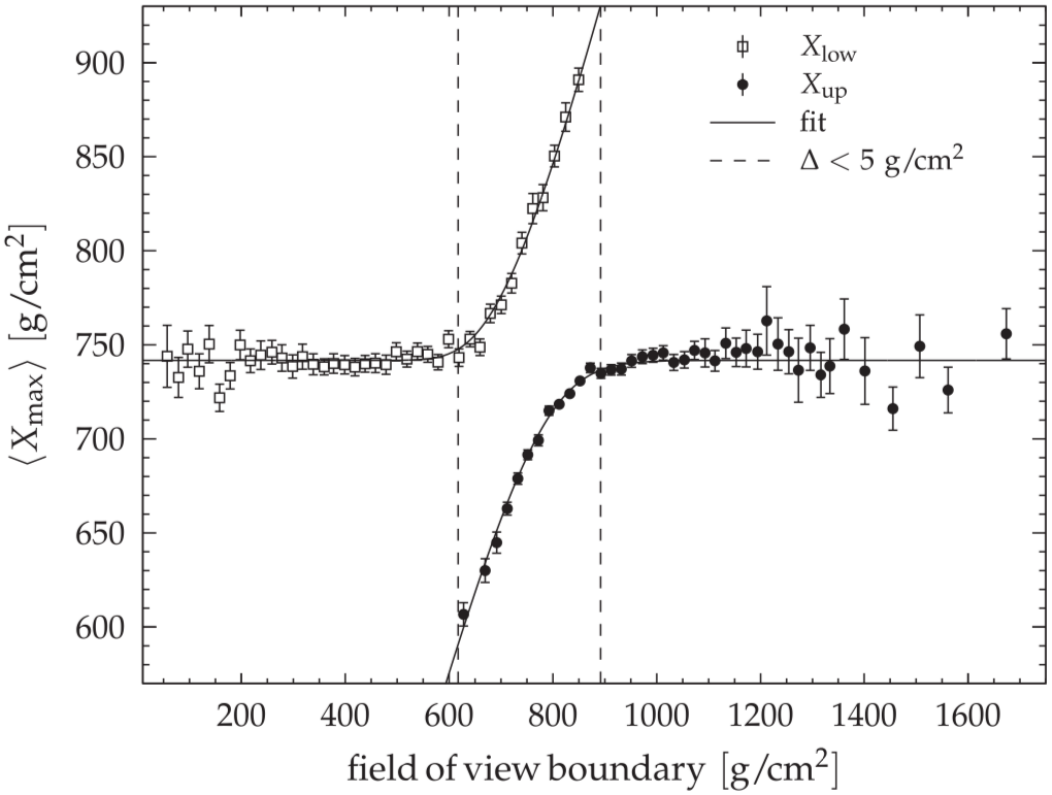


Figure 5.3. $\langle X_{max} \rangle$ as a function of the FoV limits X_{low} and X_{up} for the energy range $18.1 < \log E < 18.2$. The fiducial FoV limits (shown as dashed lines) are defined by the position that $\langle X_{max} \rangle$ deviates by more than 5 g/cm^2 from the plateaus [97].

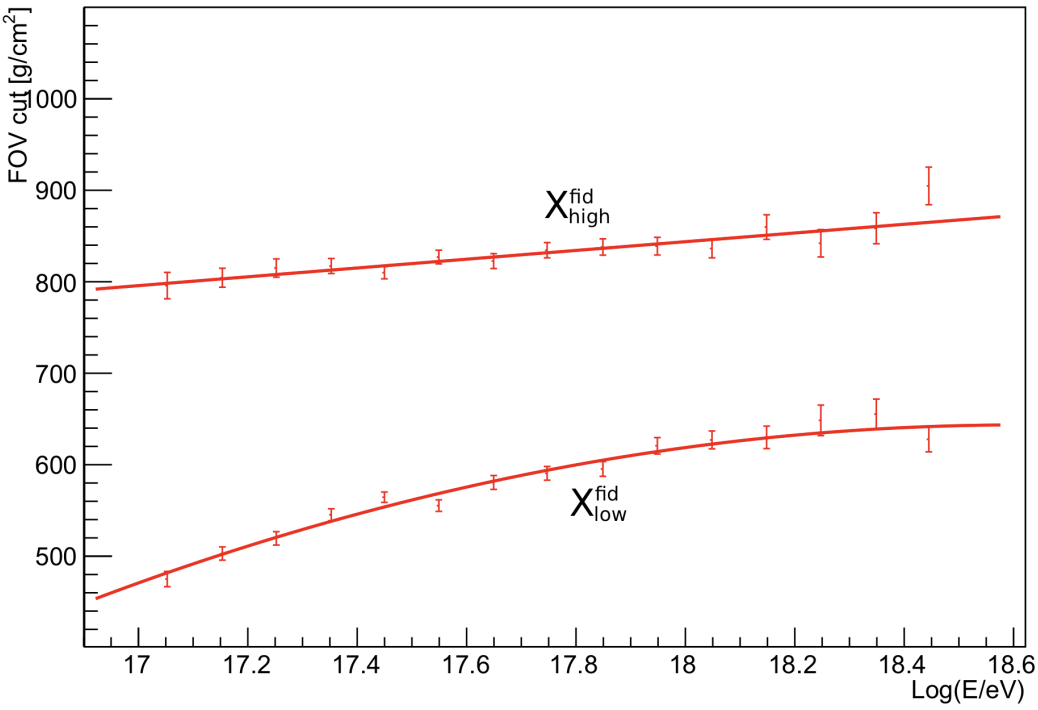


Figure 5.4. The fiducial FoV boundary limits as a function of energy.

that events with X_{max} outside the fiducial limits are retained as long as X_{low} and X_{up} pass the above conditions.

5.2.2 Detector X_{max} Acceptance

In order to determine the X_{max} detector acceptance, simulations of events with a uniform X_{max} distribution are performed. These events are folded with the detector response, processed through the reconstruction, and are subject to selection cuts. The detector acceptance as a function of X_{max} is determined based on the fraction of events surviving the full procedure. The acceptance correction is a second-order correction which is necessary due to an imperfect fiducial FoV cut. After applying the fiducial FoV cuts it is expected that the resulting reconstructed X_{max} distribution will be uniform in X_{max} between the fiducial FoV limits, since the idea of the fiducial cuts is to ensure showers with deep or shallow X_{max} are measured with equal probability. Regions outside the limits have a reduced efficiency. The relative X_{max} acceptance can be described by a flat region in-between the fiducial limits with a relative acceptance of 1, and exponential tails representing the regions of reduced acceptance outside the fiducial FoV limits. The acceptance function is given by

$$A(X_{\text{max}}) = \begin{cases} \exp\left(\frac{X_{\text{max}}-x_1}{\lambda_1}\right) & X_{\text{max}} < x_1 \\ 1 & x_1 \leq X_{\text{max}} < x_2 \\ \exp\left(-\frac{X_{\text{max}}-x_2}{\lambda_2}\right) & X_{\text{max}} \geq x_2 \end{cases} \tag{5.7}$$

where the parameters x_1, x_2, λ_1 and λ_2 are the free fit parameters. Each of these parameters are subsequently parametrised by a second-order polynomial as a function of energy

$$P_i(\log E) = A_i + B_i \log E + C_i \log^2 E \tag{5.8}$$

for $i = x_1, x_2, \lambda_1, \lambda_2$. In the case where the fit fails to converge, the parameters x_1 and x_2 can be fixed to the fiducial FoV limits. In the final analysis of the measured X_{max} distribution, events with X_{max} falling in the flat region between the fiducial FoV limits, are assigned equal weights. However, in the regions outside the limits, the events are weighted by the inverse of the relative acceptance to account for the reduced acceptance. Events in these regions have a smaller probability of being detected compared to events falling within the flat region. The re-weighting of events accounts for this non-uniform acceptance effect. This procedure is called the Λ_η method. Using the X_{max} distribution and the weights assigned by the detector acceptance, the unbiased moments of the X_{max} distribution can be then calculated.

This procedure for calculating the relative acceptance works correctly and produces the expected results for showers with energy greater than $\sim 10^{17.3}$ eV. However, it

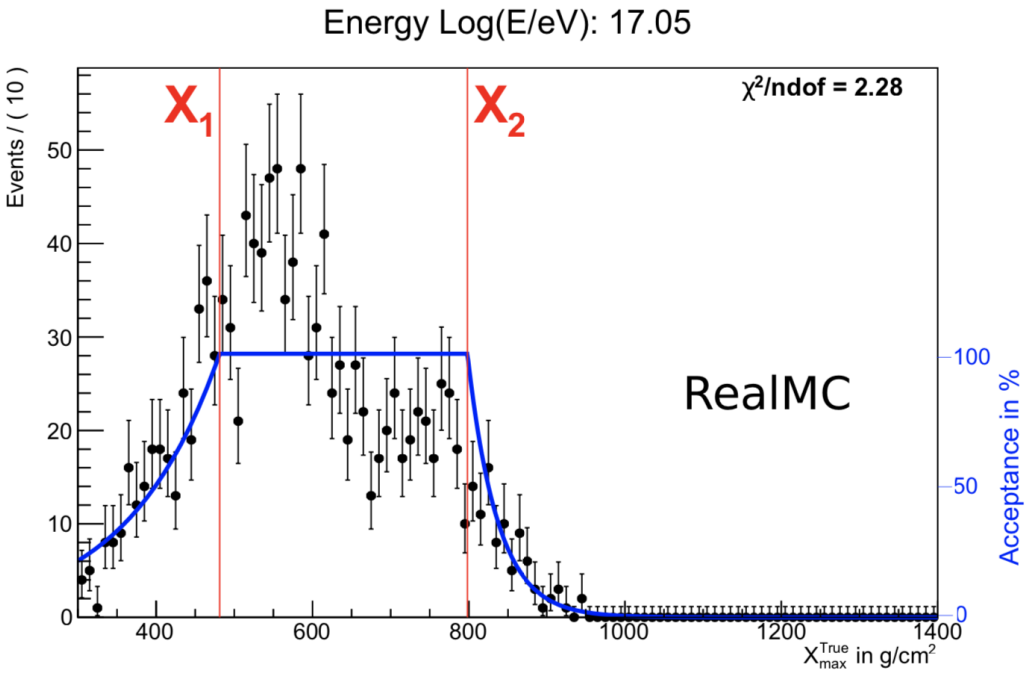
has been shown that the $X_{\max}$ detector acceptance is not flat in-between the fiducial FoV limits for low energy showers ($< 10^{17.3}$ eV) even after the FoV cuts have been applied. Two examples of the acceptance fit using Equation (5.7) are shown in Figure 5.5. The first case in Figure 5.5a shows the acceptance fit for low energy showers where Equation (5.7) does not adequately describe the detector acceptance. The cause of this effect is not clear, however, the measured HeCo events in this lower energy range are dominated by HEAT. It is possible that this highlights differences in sensitivity between the HEAT and Coihueco telescopes which are not properly being accounted for during cross-calibration. Given these issues, as well as the desire to increase the statistics at the energy limits of the analysis, a new acceptance fit must be designed in order to cope with a non-uniform acceptance. This is discussed in Section 5.3. Since the origin of the issue is unknown, the problem could be related to the calculation of the expected FoV cuts or the fiducial FoV cuts. Therefore, a conservative approach will be taken in the analysis by removing both of the FoV cuts. This is discussed further in Section 5.5, where several replacements cuts are designed in order to maintain a good quality data set.

5.3 Kernel Acceptance Model

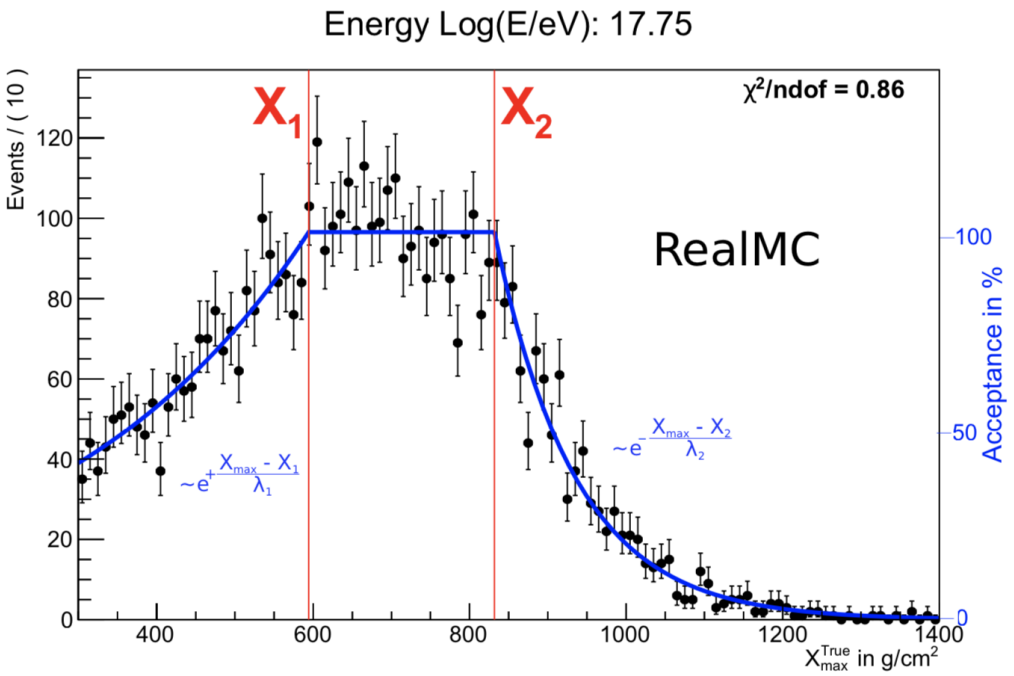
As discussed in Section 5.2.2, the $X_{\max}$ acceptance is determined using simulations of uniform $X_{\max}$ distributions. The simulated events used in this analysis are generated using CONEX showers. The showers are propagated through a detector simulation, the standard $\overline{\text{Off}}\underline{\text{line}}$ hybrid reconstruction chain, and are subject to data selection cuts. Time dependent information such as telescope calibration, atmospheric conditions and up-time fractions are taken into account in order to simulate an accurate and realistic detector state. This method of time dependent MC simulation is called RealMC. Showers are generated in two independent energy ranges for the HeCo and standard FD data sets. The parameters used in the RealMC simulations are given in Table 5.1.

Parameter	HeCo	Standard FD
Interaction model	Sibyll 2.3c	Sibyll 2.3c
Primary	p, He, N, Fe	p, He, N, Fe
Spectral Index	2.5	1.75
Energy $\log[E/eV]$	16.6 - 18.3	17.6 - 20.2
Zenith (isotropic)	0 - 80°	0 - 80°
Azimuth (flat)	0 - 360°	0 - 360°

Table 5.1. Summary of simulated events for HeCo and the Standard FD data sets.



(a)



(b)

Figure 5.5. Examples of the $X_{\max}$ relative acceptance fit using Equation (5.7) for (a) low energy showers where the fit does not describe the acceptance well, and (b) higher energy showers where the fit describes the acceptance well [192].

As discussed in Section 5.2.1.2, the fiducial FoV cuts are applied to ensure deep and shallow X_{max} values are measured with equal probability. In Section 5.2.2 it was shown that the detector acceptance is flat between the fiducial FoV limits. If the selection cuts are to be relaxed by removing the FoV cuts, the standard X_{max} acceptance function is not able to accurately model the MC data. A modification to the acceptance model is therefore required to describe the acceptance.

The X_{max} acceptance distribution without the FoV cuts does not follow a typical probability distribution. Therefore, an alternative approach must be taken to model the acceptance. The kernel estimation model is a non-parametric model used to produce an empirical probability distribution function. This method was first used in [192] for the X_{max} acceptance correction in the lowest energy bins of the HeCo data set, for which the acceptance model described in Section 5.2.2 was unable to be used. In the kernel estimation model, each of the X_{max} values in the acceptance distribution are represented by a kernel function. In this case, the kernel function is a Gaussian probability density function (PDF). A standard "fixed" kernel estimate has the form

$$\hat{f}_0(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - t_i}{h}\right) \tag{5.9}$$

where t_i are the n bins, $K(x)$ is the kernel function (Equation (5.10)) and h is the bandwidth [194].

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{5.10}$$

The kernel function spreads out the contribution from each bin in the distribution, providing a smooth estimate of the acceptance distribution. The model can be extended to an "adaptive" kernel where the bandwidth, $h \to h_i$, is allowed to vary for each data point. This gives the function flexibility to fit tall, narrow kernels to regions of high density in the distribution, and short, wide kernels for less dense regions of the distribution. This has the effect of smoothing out statistical fluctuations in the data. The choice of h is important; a value too small or too large causes under or over-fitting [194]. Expanding Equation (5.9) to an adaptive kernel estimate gives

$$\hat{f}_1(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_i} K\left(\frac{x - t_i}{h_i}\right) \tag{5.11}$$

According to [194] the best estimate of the bandwidth is given by

$$h_i^* = \rho \left(\frac{4}{3}\right)^{1/5} \sqrt{\frac{\sigma}{\hat{f}_0\left(t_i\right)}} n^{-1/5} \tag{5.12}$$

where σ is the standard deviation of the data, and $\hat{f}_0$ is the fixed kernel estimate.

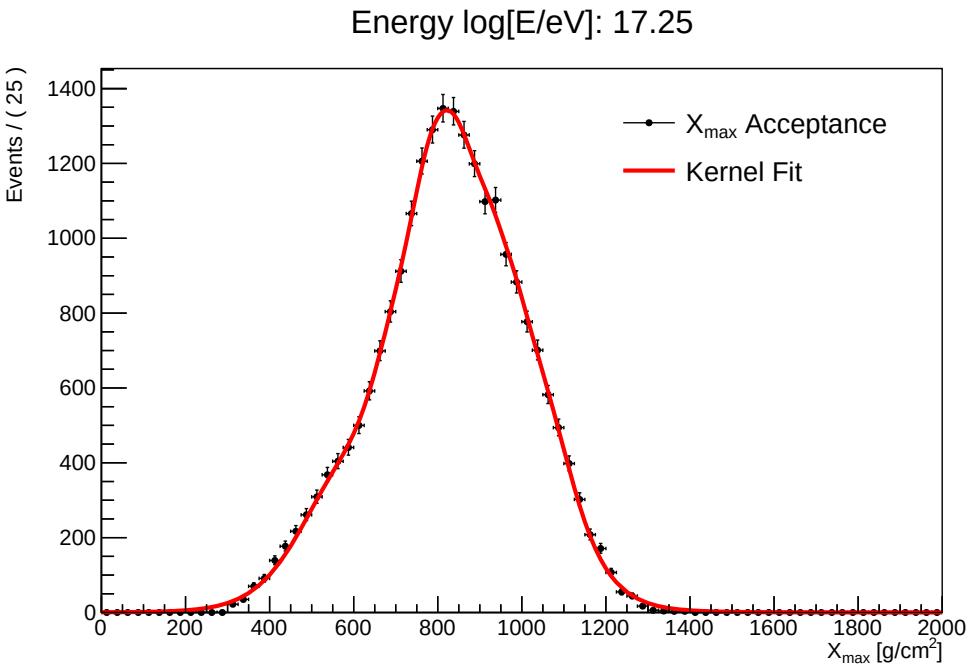


Figure 5.6. An example of the X_{max} detector acceptance derived from flat X_{max} RealMC simulations (black points) compared with the fitted kernel function (red) for the HeCo dataset in the energy range $17.2 < \log E/\text{eV} < 17.3$. The kernel function is calculated using Equation (5.11).

The kernel estimation model from the RooFit [195] package is used to obtain the empirical acceptance PDF in each energy bin. This package allows the value of ρ to be manually changed by the user in order to increase or decrease the level of smoothing applied by the kernel function. A smaller value of ρ increases the smoothing. This results in under-fitting the tails of the acceptance distribution, while increasing the value of ρ decreases the smoothing and introduces over-fitting where the kernel function varies on the level of the statistical fluctuations of the acceptance distribution. In this work $\rho = 1.5$ was chosen to obtain the best balance between under and over-fitting in the bulk and tails of the acceptance distribution. The systematic uncertainty associated with the choice of ρ is discussed in Section 5.8.4.3 and it is shown that the choice of ρ has a small, but not negligible effect on the final X_{max} moments. An example kernel fit is shown in Figure 5.6. This is repeated for all energy bins from $10^{16.8}$ to $10^{20.2}$ eV. The kernel acceptance model can therefore be used to derive the relative detector acceptance as a function of X_{max} in each energy bin.

5.4 Calculation of X_{max} Moments

The moments of the X_{max} distribution can now be calculated as described in Section 5.2.2, by weighting each X_{max} value by its relative acceptance. Without the

application of the FoV cuts, the flat region of the detector acceptance does not exist, and the entire X_{max} distribution must be weighted by the relative X_{max} acceptance rather than only the events falling outside the fiducial limits. To calculate the X_{max} moments, $\langle X_{\text{max}} \rangle$ and $\sigma_{X_{\text{max}}}$, the selected X_{max} values are directly weighted by the inverse of the relative acceptance. The X_{max} weight, $w(X_{\text{max}})$, is given by

$$w(X_{\text{max}}) = \frac{1}{\epsilon_{\text{rel}}(X_{\text{max}})} \tag{5.13}$$

where $\epsilon_{\text{rel}}(X_{\text{max}})$ is the relative acceptance derived from the kernel acceptance model. The X_{max} moments and their uncertainties can then be calculated using the weighted mean and weighted variance equations. The weighted average of a set of random variables, in this case $X_{\text{max},i}$, is given by

$$\langle X_{\text{max}} \rangle = \frac{\sum_i w_i \cdot X_{\text{max},i}}{\sum_i w_i}. \tag{5.14}$$

The unbiased weighted variance of X_{max} is given by

$$\text{Var}(X_{\text{max}}) = \frac{\left(\sum_i w_i\right)^2}{\left(\sum_i w_i\right)^2 - \sum_i w_i^2} \left(\langle X_{\text{max}}^2 \rangle - \langle X_{\text{max}} \rangle^2 \right) \tag{5.15}$$

where $\langle X_{\text{max}}^2 \rangle$ is the second non-central moment, which is given by

$$\langle X_{\text{max}}^2 \rangle = \frac{\sum_i w_i \cdot (X_{\text{max},i})^2}{\sum_i w_i} \tag{5.16}$$

The uncertainty of $\langle X_{\text{max}} \rangle$ and $\text{Var}(X_{\text{max}})$ are given by

$$\sigma^2_{\langle X_{\text{max}} \rangle} = \frac{\text{Var}(X_{\text{max}})}{N} \tag{5.17}$$

and

$$\sigma^2_{\text{Var}(X_{\text{max}})} = \frac{1}{N} \left(m_4 - \frac{N-3}{N-1} \cdot \text{Var}(X_{\text{max}})^2 \right) \tag{5.18}$$

where N is the effective sample size

$$N = \frac{\left(\sum_i w_i\right)^2}{\sum_i w_i^2}, \tag{5.19}$$

and m_4 is the fourth central moment which is given by

$$m_4 = \frac{\sum_i w_i \left(X_{\text{max},i} - \langle X_{\text{max}} \rangle\right)^4}{\sum_i w_i}. \tag{5.20}$$

In order to compare the width of the X_{max} distributions to the predictions of hadronic models, the variance must be converted into a standard deviation. The standard deviation is simply given by

$$\sigma(X_{\text{max}}) = \sqrt{\text{Var}(X_{\text{max}})}. \tag{5.21}$$

However, the uncertainty in the standard deviation of the X_{max} distribution $\sigma_{\sigma(X_{\text{max}})}$, must be obtained from error propagation through Equation (5.21). This yields

$$\sigma_{\sigma(X_{\text{max}})} = \frac{\sigma_{\text{Var}(X_{\text{max}})}}{2\sigma(X_{\text{max}})}. \tag{5.22}$$

The variance of the X_{max} distribution is artificially inflated by the detector resolution, $\sigma_{\text{res}}(X_{\text{max}})$ of the X_{max} measurements. Therefore, it must be subtracted in quadrature from the variance to obtain the true variance of the X_{max} distribution

$$\sigma^2(X_{\text{max}}) = \sigma^2(X_{\text{max}})' - \sigma^2_{\text{res}}(X_{\text{max}}) \tag{5.23}$$

where $\sigma^2(X_{\text{max}})'$ is the variance of the X_{max} distribution without the resolution correction. Finally, the uncertainty of the corrected $\sigma(X_{\text{max}})$ is obtained from error propagation to yield

$$\sigma_{\sigma(X_{\text{max}})} = \frac{\sigma(X_{\text{max}})'}{\sigma(X_{\text{max}})}\sigma'_{\sigma(X_{\text{max}})} \tag{5.24}$$

where $\sigma(X_{\text{max}})'$ and $\sigma(X_{\text{max}})$ are the uncorrected and corrected standard deviation of the X_{max} distribution, respectively, and $\sigma'_{\sigma(X_{\text{max}})}$ is the uncertainty of the uncorrected standard deviation obtained from Equation (5.22) [196].

5.5 Data Selection

It is important to remove the events which were measured either in adverse conditions or configurations in order to produce an unbiased and reliable data set. Therefore, a series of selection cuts are imposed on the data to ensure an accurate and precise reconstruction. In the standard X_{max} analysis, the FoV cuts are powerful selection tools, and are extremely important for producing unbiased X_{max} distributions. Since in this analysis, the FoV cuts are removed, replacement quality cuts must be introduced to maintain a quality dataset. Table 5.2 summarises the full list of selection cuts applied to the datasets for all X_{max} analysis in this work. The number of events passing these selection cuts is compared to the number of events passing the standard selection cuts in Figure 5.45, and discussed in Section 5.9.4. Each of the selection cuts are discussed in the following sections.

5.5.1 Hybrid Cuts

It is important to ensure the events used in the X_{max} analysis are of high quality, with a reliable estimate of the shower geometry. These events are selected using the *IsHybrid* cut. Limiting the data set to only hybrid events (reconstructed using information from both the FD and SD) provides the most accurate geometry as discussed in Section 3.5.3. The *IsCLF* and *IsXLF* cuts are set to remove any events which were generated by the laser facilities. The standard X_{max} analysis cuts reject events if the hybrid station is not a time-over-threshold (ToT) trigger. This can be relaxed to include threshold (Th) triggered stations if they are selected by the hybrid reconstruction algorithm. While the requirement of a ToT trigger avoids events with SD stations triggering on atmospheric background muons, the chance of the hybrid station triggering on atmospheric background muons both in time-coincidence and spatial-coincidence with the hybrid geometry fit is relatively low. An additional cut of 1500 m on the distance between the shower core position and the SD station used during hybrid reconstruction is applied to reject events which have a reconstructed core position too far from the hybrid station. Such events are unreliable.

5.5.2 SD Trigger Probability

The standard trigger probability cut removes events where the combined trigger probability of all stations

$$P = 1 - \prod_i (1 - P_i) \tag{5.25}$$

is less than 0.9. Here P_i is the Lateral Trigger Probability (LTP) of station i based on LTP parametrisations derived from simulations. An additional cut removes the event if the difference between the proton and iron trigger probability is greater than 0.05 to ensure no composition bias is introduced by the SD trigger.

These two selection cuts are replaced by the data driven approach used to derive LTP parametrisations in Chapter 4. This new cut is called *minPBrassData*. The new cut still uses Equation (5.25) with a requirement that the trigger probability is less than 0.9, however, the value of P_i is determined using the LTP parametrisations from Chapter 4. While the LTP parametrisations derived from simulations require a choice of mass composition, the parametrisations derived in Chapter 4 using hybrid data naturally account for the true mass composition. Furthermore, it was shown in Chapter 4 that the SD 750 m array is fully efficient for a single station using the new triggers down to $10^{16.7}$ eV for the HeCo dataset, and to $10^{17.4}$ eV for the standard FD dataset which is limited to the trigger efficiency of the 1500 m array. However, prior to 2014, the new triggers had not been implemented, therefore for data prior to this time, the LTP parametrisation for only the old triggers are used to determine

the trigger probability. The parametrisations have been implemented in the Auger Offline framework as an ADST selection cut. Note that the energy threshold for full efficiency for the old triggers is $10^{16.9}$ eV and $10^{17.55}$ eV for the 750 m array and the 1500 m array, respectively.

5.5.3 HeCo Cuts

In order to separate the HeCo and standard FD data sets, several cuts are used to accept or reject events which fall into each of the two categories. For the HeCo data set, only events with an energy below $10^{18.1}$ eV are accepted using the *keepHECOor-CoihecoHEAT* cut. Higher energy events are assigned to the standard FD data set. In addition to this cut, the HeCo data set rejects any events which were recorded by an FD other than HEAT or Coihueco using the *eyeCut* cut. Finally, the *heatOrientationUp* cut ensures HEAT was operating in its upward configuration.

5.5.4 Data Acquisition Cuts

There are time periods for which the FD measurements are unreliable. This can be due to missing calibration information or malfunctions for example. Events during these bad periods are rejected using the *badFDPeriodRejection* cut. The geometry reconstruction can be sensitive to the set of included pixels. Therefore, it is important to remove events which contain issues with photomultiplier tubes (PMTs) used in the reconstruction process. The *skipSaturated* and*noBadPixelsInPulse* cuts are applied to remove events with contain saturated PMTs or PMTs with missing calibration information. A correction is applied to the data to account of imprecision in the GPS timing at the nano-second level [197]. The *good10MHzCorrection* cut is applied to reject events without such a correction.

5.5.5 Atmospheric Cuts

The *hasMieDatabase* cut ensures events have access to the measurements of the aerosol content of the atmosphere as measured by the Central Laser Facility (CLF) and eXtreme Laser Facility (XLF). The *maxVAOD* rejects events which have a vertical aerosol optical depth (VAOD) greater than 0.1 at the reference height of 3 km above ground level. This avoids large systematic uncertainties associated with the reconstructed energy but maintains a large data sample. Finally, the cloud cuts are applied to remove events with a large cloud fraction based on measurements from the cloud cameras and Lidars. Events affected by cloud are susceptible to large biases in their reconstructed energy due to scattering and absorption of the fluorescence and Cherenkov light by clouds.

5.5.6 Reconstruction Cuts

Both of the expected FoV cut and the fiducial FoV cut have been removed since they are not producing flat acceptance, as described in Section 5.2.2. This also has the effect of increasing the number of events in the lowest and highest energy bins. Instead the acceptance is corrected using the kernel model to weight each value of X_{max} according to its corresponding relative acceptance. However, it is important to maintain a good quality data set to ensure the correct weighting is applied to each event in the calculation of the X_{max} moments. Therefore, a series of replacement cuts have been applied to the data. The *xMaxInFOV* cut was added to ensure the reconstructed X_{max} is inside the geometrical FoV of the FD telescope. The reconstructed X_{max} can be significantly biased for events not passing this cut. Additionally, the uncertainty in the reconstructed X_{max} is required to be less than the reconstructed value of X_{max} itself as a sanity cut to remove very poorly reconstructed events.

The expected FoV cut in the standard analysis removes events with a viewing angle of 20° or less. This is important since events with smaller viewing angles are likely contaminated by a large fraction of Cherenkov light. Such events are highly susceptible to systematic errors in the geometry reconstruction. For this reason, the viewing angle of events in this analysis is required to be greater than 20°. Another effect of the expected FoV cut is to remove events with an expected X_{max} error larger than 40 g/cm^2. To ensure good quality events, the reconstructed X_{max} uncertainty is required to be less than 40 g/cm^2 as a replacement for the expected FoV cut. In addition to these cuts the uncertainty in the reconstructed energy is required to be less than 20%.

The *depthTrackLength* cut rejects events which have short tracks across the FD camera. Such events are prone to systematic uncertainties in the geometry reconstruction and poor profile fits. The cut is applied to the atmospheric depth viewed to account for different shower geometries. Only events with a track length greater than 200 g/cm^2 are retained. This value is chosen to balance between removing events at low energy and maintaining long tracks. For the HeCo data set, the cut is removed in order to retain a larger data set at low energy. The *maxDepthHole* requires that any gaps in the energy deposit profile are less than 20% of the total for the standard FD data set. This is increased to 30% for the HeCo data set. Such gaps can arise from the gaps between FD cameras of separate telescopes.

Finally, the *profileChi2Sigma* cut is applied to ensure the shower profile fit is reliable. The cut rejects outliers in the χ^2 distribution based on the normalised χ^2

$$z = \frac{\chi^2 - Ndf}{\sqrt{2 \cdot Ndf}} \tag{5.26}$$

where Ndf is the number of degrees of freedom in the profile fit. According to the central limit theorem, z is expected to be normally distributed. Since the profile fit is obtained using a maximum likelihood method, the distribution is not centred at 0. Therefore, a correction of -1.42 and cut of 5 is applied to Equation (5.26) for the HeCo and a correction of -1.1 and cut of 3 for the standard FD data sets, respectively.

5.5.7 Stereo Events

It is possible that an event is detected by more than one FD, especially at higher energies in the standard FD data set. Each FD event is reconstructed with varying quality. It is therefore important to average these events with weights corresponding to the uncertainty in their reconstructed X_{max} and energy. Therefore, the energy and X_{max} of the events are weighted by the inverse of their variances:

$$\sigma^2(E) = \left[\sum_i \frac{1}{\sigma_i^2(E)}\right]^{-1} \tag{5.27}$$

$$E = \sigma^2(E) \sum_i \frac{E_i}{\sigma_i^2(E)} \tag{5.28}$$

$$\sigma^2(X_{\text{max}}) = \left[\sum_i \frac{1}{\sigma_i^2(X_{\text{max}})}\right]^{-1} \tag{5.29}$$

$$X_{\text{max}} = \sigma^2(X_{\text{max}}) \sum_i \frac{X_{\text{max}}^i}{\sigma_i^2(X_{\text{max}})} \tag{5.30}$$

where $\sigma_i(E)$ and $\sigma_i(X_{\text{max}})$ are the estimated uncertainties of E_i and X_{max}^i from the ith FD.

This procedure is only done for the standard FD data set. The HeCo data set uses events measured only by HEAT, Coihueco or the virtual FD HeCo with a preference for the full virtual FD measurement where it is available.

Cut name	HeCo	Standard FD
IsCLF	False	False
IsXLF	False	False
IsHybrid	True	True
keepHECOorCoihecoHEAT	18.1	18.1
eyeCut	100000	1111
heatOrientationUp	True	-
badFDPeriodRejection	True	True
skipSaturated	True	True
noBadPixelsInPulse	True	True
good10MHzCorrection	True	True
hasMieDatabase	True	True
maxVAOD	0.1	0.1
cloudCutXmaxPRD14	21, -10.5, 10.5	21, -10.5, 10.5
maxCoreTankDist	1500	1500
maxZenithFD	85	90
minLgEnergyFD	16.7	17.8
minPBrassData	0.9	0.9
depthTrackLength	-	200
XmaxErrorLessThenXmax	True	True
xMaxInFOV	True	True
minViewAngle	20	20
xMaxError	40	40
energyError	0.2	0.2
maxDepthHole	30	20
profileChi2Sigma	5.0, -1.42	3.0, -1.1

Table 5.2. List of selection cuts for HeCo and standard FD datasets with corresponding cut values.

5.6 Reconstruction Bias and Detector Resolution Corrections

In order to evaluate the reconstruction bias and detector resolution, the differences between the simulated and reconstructed calorimetric energy and X_{max} are studied. The biases and detector resolution are parametrised as a function of energy. These parametrisations can then be used to correct the moments of the X_{max} distributions for each composition. While this is relatively simple for simulated data with a known composition, the data cannot be corrected in the same way without first determining composition fractions which are compatible with the data. This is discussed in Section 5.6.1 and Section 5.8.

5.6.1 Fraction Fits

In order to estimate the reconstruction bias and X_{max} resolution for the data, an estimate of the composition of the data must be obtained. An exponentially modified Gaussian (EMG) distribution describes the observed X_{max} distributions well [198]. This information can be used to fit a functional form to the simulated X_{max} distributions in a given energy bin. The EMG variable, Z can be expressed as $Z = X + Y$ where X and Y are independent; X is a Gaussian with mean μ and variance σ^2, and Y is exponential with rate λ. The PDF of the EMG distribution is derived via a convolution of the Gaussian and exponential PDFs, and can be written as

$$G(x) = \frac{1}{2\lambda} \exp\left(\frac{\mu - x}{\lambda} + \frac{\sigma^2}{2\lambda^2}\right) \mathrm{erfc}\left(\frac{\mu - x + \sigma^2/\lambda}{\sigma\sqrt{2}}\right) \tag{5.31}$$

where erfc is the complementary error function, defined as

$$\mathrm{erfc}(x) = 1 - \mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt \tag{5.32}$$

The goal of these fractions fits is to determine the fraction of proton, helium, nitrogen and iron MC events in each energy bin which best matches the reconstructed X_{max} distributions. To this end, an EMG function is fit to the MC events for each composition and energy bin. Examples of these fits are shown in Figure 5.7. Following the individual composition fits, the data X_{max} distributions are fit to the weighted sum of the four composition EMG functions with the normalisation (fraction) of the EMG functions as free parameters. This can be written as

$$G(x) = f_{\mathrm{p}} G_{\mathrm{p}}(x) + f_{\mathrm{He}} G_{\mathrm{He}}(x) + f_{\mathrm{N}} G_{\mathrm{N}}(x) + f_{\mathrm{Fe}} G_{\mathrm{Fe}}(x) \tag{5.33}$$

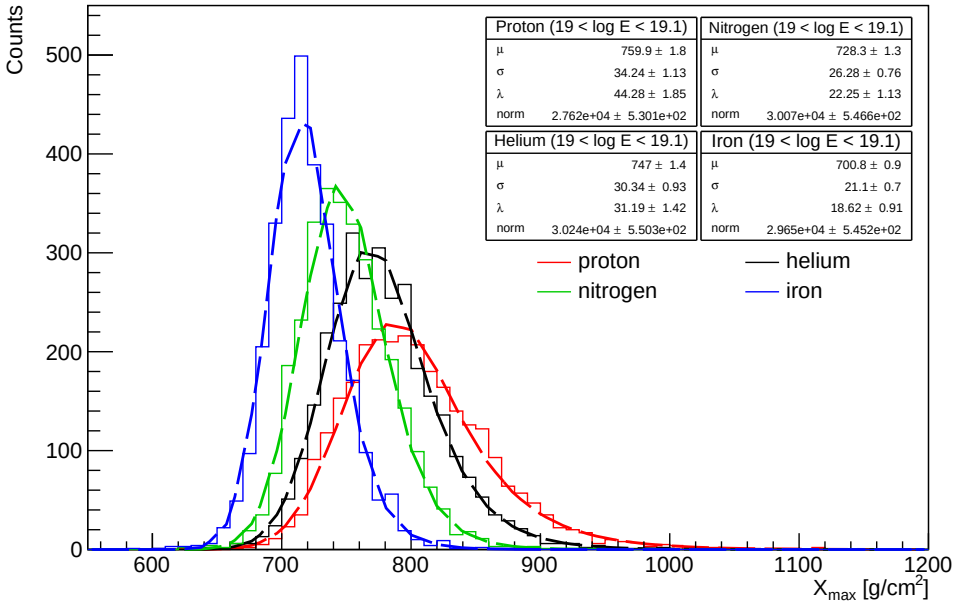


Figure 5.7. Example of the EMG fits (dashed lines) compared to the MC X_{max} distributions (solid lines) for proton (red), helium (black), nitrogen (green), and iron (blue) in the energy range $19.0 < \log E/\mathrm{eV} < 19.1$. The best fit EMG parameters and their uncertainties are included.

where f_{p}, f_{He}, f_{N} and f_{Fe} are the composition fractions for proton, helium, nitrogen and iron, respectively. The number of free parameters can be reduced to three by imposing the requirement that the sum of the four fractions is equal to 1. The four compositions fractions can therefore be written as

$$f_{\mathrm{p}} = \eta_1 \tag{5.34}$$

$$f_{\mathrm{He}} = (1 - \eta_1)\,\eta_2 \tag{5.35}$$

$$f_{\mathrm{N}} = (1 - \eta_1)\,(1 - \eta_2)\,\eta_3 \tag{5.36}$$

$$f_{\mathrm{Fe}} = 1 - f_{\mathrm{p}} - f_{\mathrm{He}} - f_{\mathrm{N}} \tag{5.37}$$

where η_1, η_2 and η_3 are the three free fit parameters in the fraction fits. An example of a MC X_{max} distribution with fitted composition fractions is compared to the data X_{max} distribution in Figure 5.8. The agreement in the shape and the mean and standard deviation of the distribution is good. Therefore, these composition fractions will be used in the following sections to estimate the reconstruction bias and X_{max} resolution of the data.

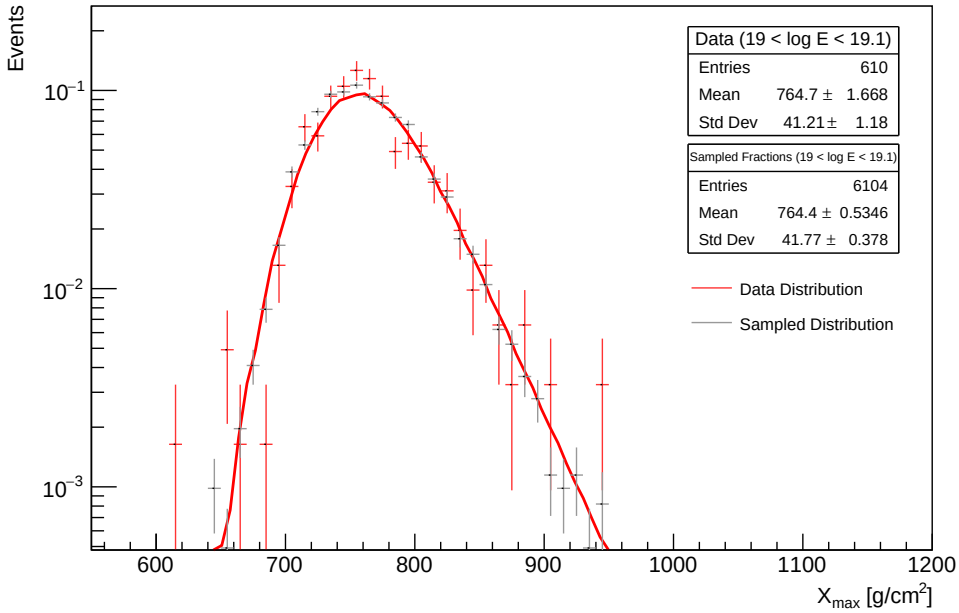


Figure 5.8. Example comparison of the data X_{max} distribution (red) and an X_{max} distribution sampled with weights of each composition (proton, helium, nitrogen and iron) corresponding to the fitted fractions (grey). Also shown as the solid red line is a weighted (by the fractions) sum of the four EMG functions shown in Figure 5.7. The same energy range of $19.0 < \log E/\mathrm{eV} < 19.1$ from Figure 5.7 is used here. The agreement between the data and sampled X_{max} distributions is good.

5.6.2 Energy Reconstruction Bias

The energy bias of simulated events is defined as the relative bias between the reconstructed and true calorimetric energy ($E^{\text{cal}}_{\text{rec}}$ and $E^{\text{cal}}_{\text{true}}$, respectively). This can be expressed as

$$\Delta E = \frac{E^{\text{cal}}_{\text{rec}} - E^{\text{cal}}_{\text{true}}}{E^{\text{cal}}_{\text{rec}}} \tag{5.38}$$

The mean relative energy bias, $\langle \Delta E \rangle$ as a function of $\log E^{\text{cal}}_{\text{rec}}$ is shown for the HeCo and standard FD data sets in Figures 5.9a and 5.9b, respectively, for four different compositions: pure proton, pure iron, a 50% proton/iron mix and the data mix. The fraction fits from Section 5.6.1 are used to sample MC X_{max} distributions with composition weights corresponding to the fitted fractions. To correct for this bias, the mean relative energy bias is parametrised as a function of $\log E^{\text{cal}}_{\text{rec}}$, using a second-order polynomial. The best fit parametrisations are also shown in Figures 5.9a and 5.9b. The fit function has the form

$$\langle \Delta E \rangle (E^{\text{cal}}_{\text{rec}}/\text{eV}) = A + B \log \left(E^{\text{cal}}_{\text{rec}}/\text{eV} \right) + C \log^2 \left(E^{\text{cal}}_{\text{rec}}/\text{eV} \right) \tag{5.39}$$

where the best fit parameters A, B and C are given for each of the fitted compositions in Table 5.3 for both the HeCo and standard FD data sets. The reconstructed calorimetric energy can then be corrected by

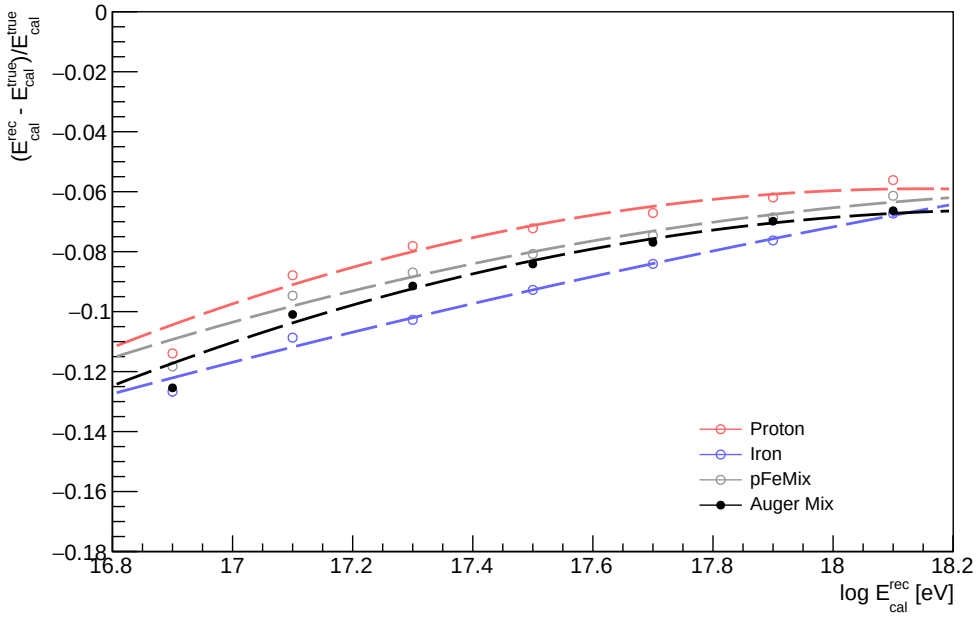
$$E^{\text{cal}}_{\text{true}} = E^{\text{cal}}_{\text{rec}} \left(1 - \langle \Delta E \rangle \right) \tag{5.40}$$

and the total energy uses the same relative energy bias correction

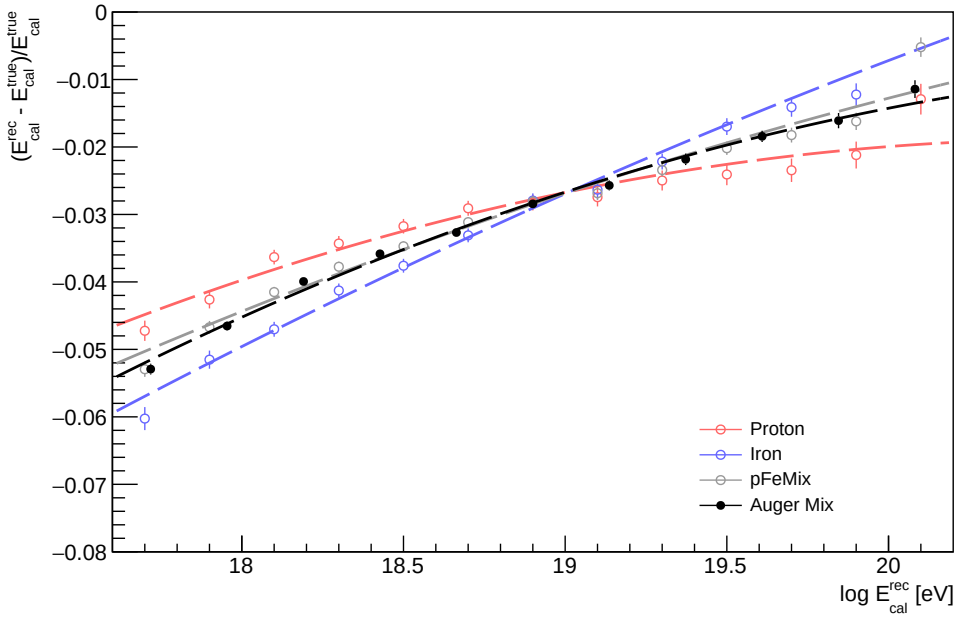
$$E^{\text{total}}_{\text{true}} = E^{\text{total}}_{\text{rec}} \left(1 - \langle \Delta E \rangle \right) \tag{5.41}$$

Dataset	Composition	A	B	C
HeCo	Proton	-9.63 ± 0.85	1.05 ± 0.097	-0.0291 ± 0.0028
	Iron	-2.79 ± 0.94	0.263 ± 0.11	-0.00621 ± 0.003
	pFeMix	-6.12 ± 0.63	0.653 ± 0.072	-0.0176 ± 0.0021
	Auger Mix	-8.67 ± 0.52	0.94 ± 0.059	-0.0257 ± 0.0017
FD	Proton	-1.33 ± 0.31	0.127 ± 0.033	-0.00309 ± 0.00087
	Iron	-0.956 ± 0.29	0.0766 ± 0.031	-0.00146 ± 0.00081
	pFeMix	-0.944 ± 0.21	0.0807 ± 0.022	-0.00171 ± 0.00059
	Auger Mix	-1.42 ± 0.18	0.132 ± 0.019	-0.00305 ± 0.00052

Table 5.3. The best fit energy bias parameters for each composition mix for both the HeCo and the standard FD data sets.



(a)



(b)

Figure 5.9. The mean relative energy bias as a function of $\log E_{\mathrm{rec}}^{\mathrm{cal}}$ for (a) the HeCo dataset, and (b) the standard FD dataset for pure proton (red), pure iron (blue), 50% proton/iron mix (grey) and the fitted Auger mix (black) from Section 5.6.1. Also shown as the dashed lines are the best fit parametrisation based on Equation (5.39) for each composition and the data.

5.6.3 X_{max} Reconstruction Bias

In a similar way to the energy reconstruction bias, the X_{max} reconstruction bias is defined as the difference between the reconstructed and true X_{max} ($X_{\text{max}}^{\text{rec}}$ and $X_{\text{max}}^{\text{true}}$, respectively). This can be expressed as

$$\Delta X_{\text{max}} = X_{\text{max}}^{\text{rec}} - X_{\text{max}}^{\text{true}} \tag{5.42}$$

Since the reconstructed calorimetric energy has already been corrected using the parametrisation of the mean relative energy bias from Section 5.6.2, the mean X_{max} bias, $\langle \Delta X_{\text{max}} \rangle$ is instead parametrised as a function of $\log E_{\text{true}}^{\text{cal}}$, using a second-order polynomial of the form

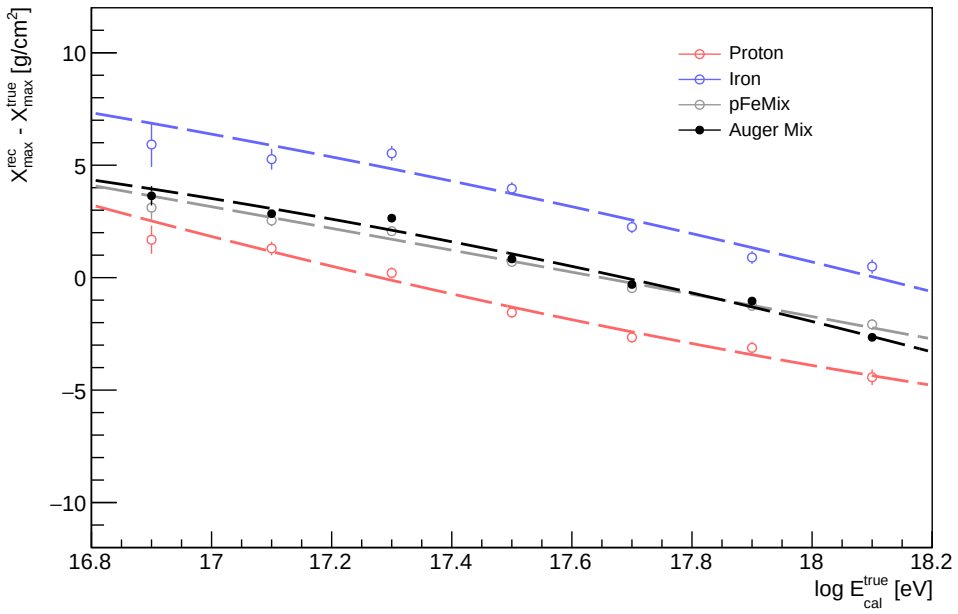
$$\langle \Delta X_{\text{max}} \rangle (E_{\text{true}}^{\text{cal}}/\text{eV}) = A + B \log \left(E_{\text{true}}^{\text{cal}}/\text{eV} \right) + C \log^2 \left(E_{\text{true}}^{\text{cal}}/\text{eV} \right) \tag{5.43}$$

The mean X_{max} bias with the best fit parametrisations for each composition (pure proton, pure iron, 50% proton/iron mix and the fitted Auger mix from Section 5.6.1) is shown in Figures 5.10a and 5.10b for the HeCo and standard FD data sets, respectively. The best fit parameters of Equation (5.43) are shown in Table 5.4. The reconstructed X_{max} is corrected by

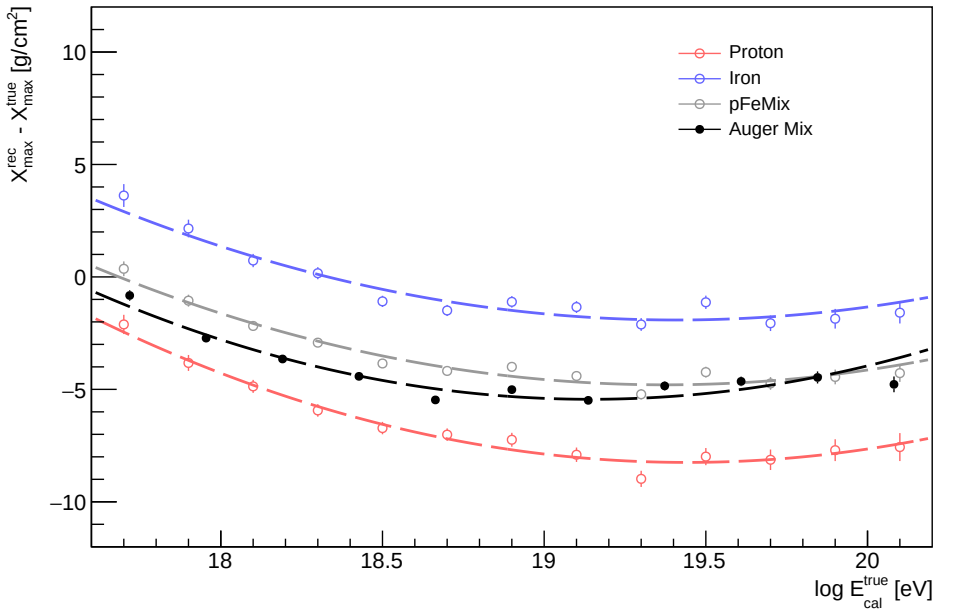
$$X_{\text{max}}^{\text{true}} = X_{\text{max}}^{\text{rec}} - \langle \Delta X_{\text{max}} \rangle \tag{5.44}$$

Dataset	Composition	A	B	C
HeCo	Proton	431 ± 300	-43.7 ± 35	1.08 ± 0.99
	Iron	-134 ± 370	21.4 ± 42	-0.775 ± 1.2
	pFeMix	57.9 ± 230	-1.65 ± 27	-0.0922 ± 0.76
	Auger Mix	-240 ± 180	33 ± 21	-1.1 ± 0.59
FD	Proton	714 ± 81	-74.3 ± 8.6	1.91 ± 0.23
	Iron	621 ± 76	-64.2 ± 8	1.65 ± 0.21
	pFeMix	626 ± 56	-65.2 ± 5.9	1.68 ± 0.16
	Auger Mix	737 ± 48	-77.6 ± 5.1	2.03 ± 0.14

Table 5.4. The best fit X_{max} bias parameters for each composition mix for both the HeCo and the standard FD data sets.



(a)



(b)

Figure 5.10. The mean $X_{\max}$ bias as a function of $\log E_{\text{true}}^{\text{cal}}$ for (a) the HeCo dataset, and (b) the standard FD dataset for pure proton (red), pure iron (blue), 50% proton/iron mix (grey) and the fitted Auger mix (black) from Section 5.6.1. Also shown as the dashed lines are the best fit parametrisation based on Equation (5.43) for each composition.

5.6.4 X_{max} Resolution

In addition to the mean X_{max} bias correction, the detector resolution, σ_{res} can be derived from the ΔX_{max} distribution at a given energy since

$$\sigma_{\text{res}} = \sigma(\Delta X_{\text{max}}) \tag{5.45}$$

where $\sigma(\Delta X_{\text{max}})$ is the standard deviation of the ΔX_{max} distribution. This is independent of any intrinsic shower to shower fluctuations of showers, and represents the pure X_{max} detector resolution. Therefore, σ_{res} can be used to correct the variance of the measured X_{max} distributions and remove the detector resolution effects. The detector resolution is parametrised as a function of $\log E^{\text{cal}}_{\text{true}}$ using a second-order polynomial of the form

$$\sigma_{\text{res}}(E^{\text{cal}}_{\text{true}}/\text{eV}) = A + B\log\left(E^{\text{cal}}_{\text{true}}/\text{eV}\right) + C\log^2\left(E^{\text{cal}}_{\text{true}}/\text{eV}\right) \tag{5.46}$$

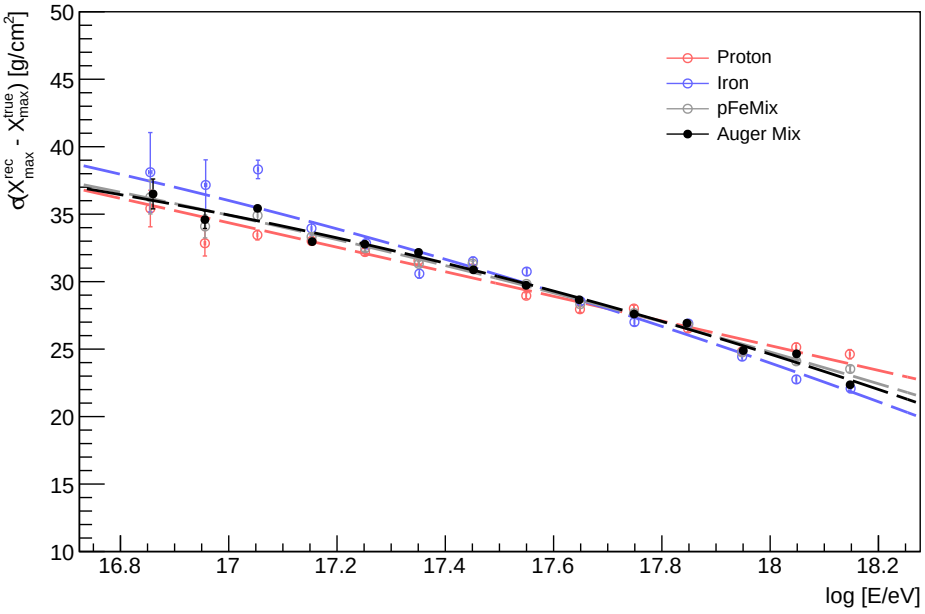
The detector resolution with the best fit parametrisations for each composition (pure proton, pure iron, 50% proton/iron mix and the fitted Auger mix from Section 5.6.1) is shown in Figures 5.11a and 5.11b for the HeCo and standard FD data sets, respectively. The best fit parameters of Equation (5.46) are shown in Table 5.5. The reconstructed variance of the X_{max} distributions, $\sigma^2(X^{\text{rec}}_{\text{max}})$ is then corrected by subtracting the detector resolution in quadrature

$$\sigma^2(X^{\text{true}}_{\text{max}}) = \sigma^2(X^{\text{rec}}_{\text{max}}) - \sigma^2_{\text{res}} \tag{5.47}$$

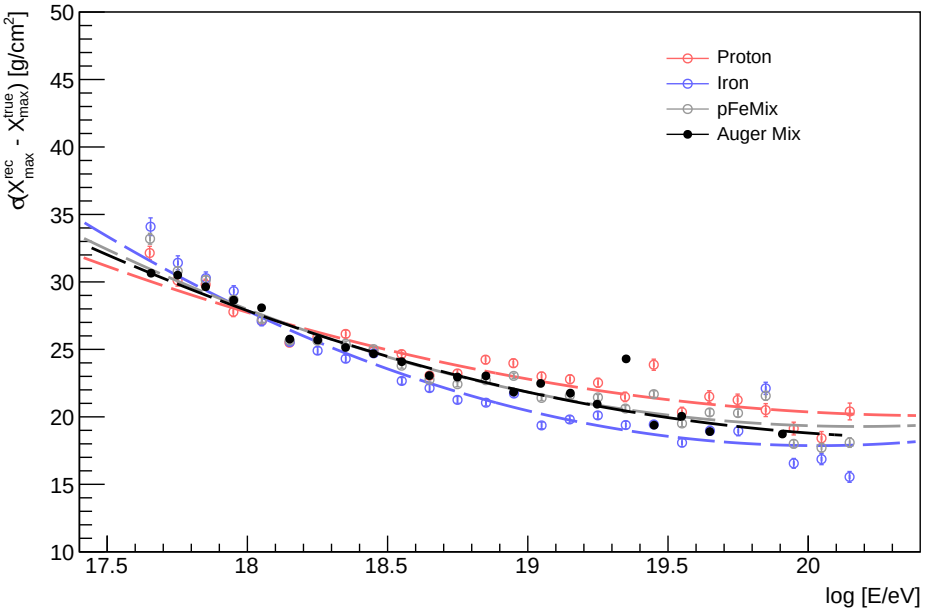
as described in Section 5.4.

Dataset	Composition	A	B	C
HeCo	Proton	104 ± 1.1	0.546 ± 0.064	-0.275 ± 0.0034
	Iron	-347 ± 1.3	55.2 ± 0.076	-1.92 ± 0.004
	pFeMix	-208 ± 0.81	37.4 ± 0.049	-1.36 ± 0.0026
	Auger Mix	-494 ± 0.63	70.2 ± 0.038	-2.3 ± 0.002
FD	Proton	545 ± 0.82	-51.3 ± 0.053	1.25 ± 0.0023
	Iron	987 ± 0.74	-96.7 ± 0.047	2.41 ± 0.0021
	pFeMix	763 ± 0.56	-73.7 ± 0.036	1.83 ± 0.0016
	Auger Mix	650 ± 8.4	-61.5 ± 0.89	1.5 ± 0.024

Table 5.5. The best fit X_{max} detector resolution parameters for each composition mix for both the HeCo and the standard FD data sets.



(a)



(b)

Figure 5.11. The detector $X_{\max}$ resolution as a function of $\log E_{\text{true}}^{\text{cal}}$ for (a) the HeCo dataset, and (b) the standard FD dataset for pure proton (red), pure iron (blue), 50% proton/iron mix (grey) and the fitted Auger mix (black) from Section 5.6.1. Also shown as the dashed lines are the best fit parametrisation based on Equation (5.46) for each composition.

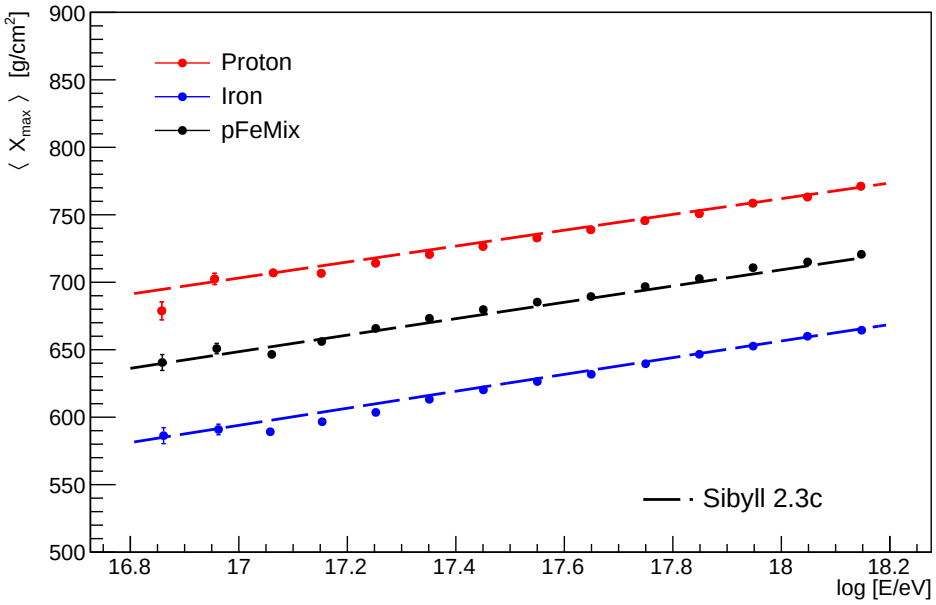
5.7 End-to-End Monte Carlo Analysis

In order to validate the modified selection cuts and the kernel acceptance model, an end-to-end MC check of the analysis procedure has been performed. Simulated events were generated in two separate energy ranges between $10^{16.8}$ to $10^{18.3}$ eV for the HeCo data set and $10^{17.8}$ to $10^{20.2}$ eV for the standard FD data set. The MC data are simulated using the Sibyll 2.3c hadronic model [81] for a mix of 25% each of proton, helium, nitrogen and iron. These simulated data sets with known composition are processed through the full reconstruction and analysis sequence. The moments of the reconstructed distributions are then compared with predications of the Sibyll 2.3c hadronic model for three different scenarios, namely pure proton, pure iron and a mixed composition consisting of 50% proton and 50% iron.

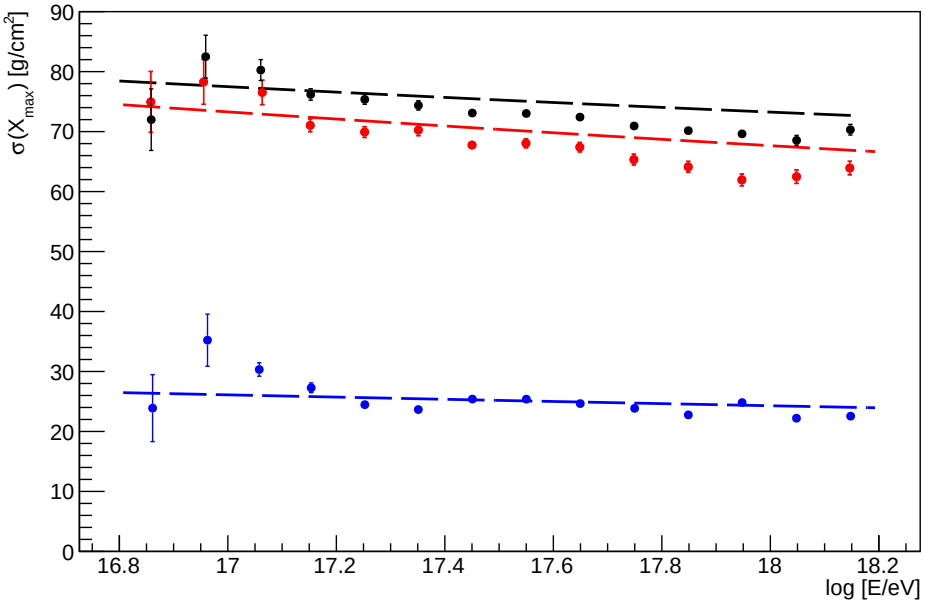
5.7.1 Monte Carlo X_{max} Moments

The first two moments of the reconstructed MC X_{max} distributions are calculated as described in Section 5.4, including the X_{max} acceptance weighting using the kernel acceptance fits described in Section 5.3. The reconstruction biases in energy and X_{max} are fully corrected, and the detector resolution is subtracted from $\sigma\left(X_{\text{max}}\right)$ in quadrature for each composition scenario using the parametrisations determined in Section 5.6. The calculated X_{max} moments, $\langle X_{\text{max}} \rangle$ and $\sigma\left(X_{\text{max}}\right)$ are compared to the predictions from the simulated hadronic model; Sibyll 2.3c. The comparisons are shown for both the HeCo data set in Figures 5.12a and 5.12b and the standard FD data set in Figures 5.13a and 5.13b.

There is good agreement between the calculated $\langle X_{\text{max}} \rangle$, however a significant bias is present in $\sigma\left(X_{\text{max}}\right)$, where the calculated value of $\sigma\left(X_{\text{max}}\right)$ is progressively underestimated with increasing energy. It also appears that the bias is composition dependent. It is not likely that there is an inherent composition dependence, but instead a dependence on the width of the X_{max} distribution, which decreases with mass. This bias was investigated in detail by comparing $\sigma\left(X_{\text{max}}\right)$ as calculated using the simulated (true) X_{max} values and the reconstructed X_{max} values. An example of this is shown in Figure 5.14 for pure proton composition in the HeCo data set. The detector resolution has not been subtracted from the reconstructed $\sigma\left(X_{\text{max}}\right)$. Note that the reconstructed $\sigma\left(X_{\text{max}}\right)$ is larger than the true $\sigma\left(X_{\text{max}}\right)$ at low energy as expected, however the reconstructed $\sigma\left(X_{\text{max}}\right)$ becomes increasingly closer to the true $\sigma\left(X_{\text{max}}\right)$ at higher energy. Once the detector resolution is subtracted in quadrature from the reconstructed $\sigma\left(X_{\text{max}}\right)$, the reconstructed $\sigma\left(X_{\text{max}}\right)$ will become less than the true $\sigma\left(X_{\text{max}}\right)$ at higher energy, revealing the bias observed in Figures 5.12b and 5.13b. Since the bias is only observed once the simulated data set is processed

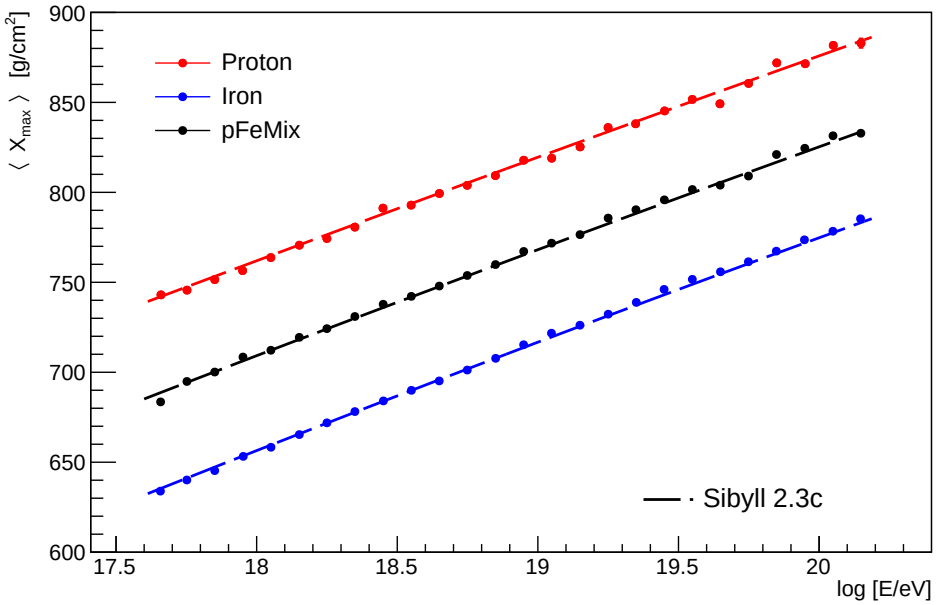


(a)

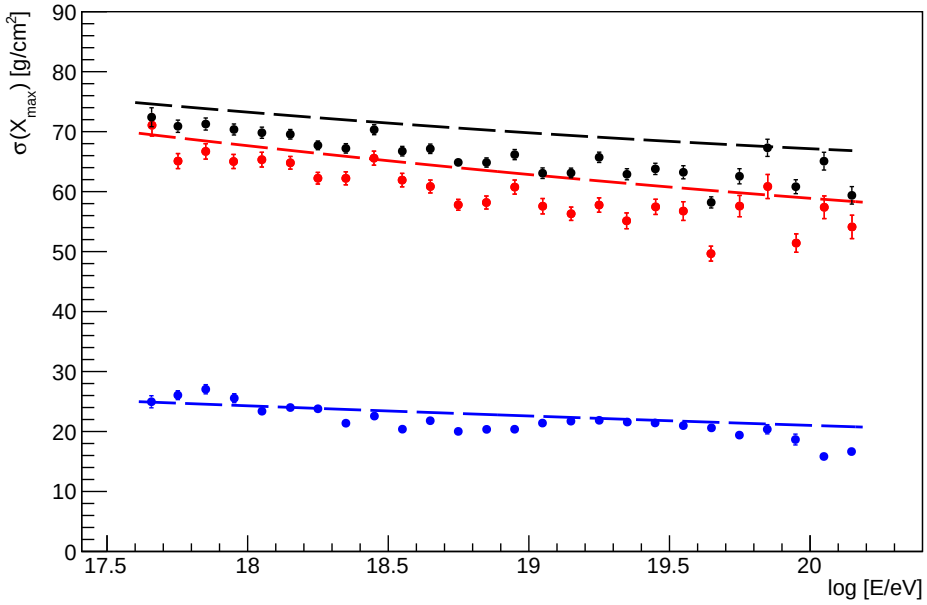


(b)

Figure 5.12. The moments of the MC $X_{\max}$ distributions from the HeCo data set for pure proton (red), pure iron (blue) and a mix of 50% each of proton and iron (black). The moments are compared to the prediction of the simulated hadronic model, Sibyll 2.3c (dashed lines). There is good agreement between the predicted mean (a) and the prediction of Sibyll 2.3c, however there is a clear energy dependent bias in $\sigma\left(X_{\max}\right)$ (b) compared with the prediction of Sibyll 2.3c.



(a)



(b)

Figure 5.13. The moments of the MC $X_{\max}$ distributions from the standard FD data set for pure proton (red), pure iron (blue) and a mix of 50% each of proton and iron (black). The moments are compared to the prediction of the simulated hadronic model, Sibyll 2.3c (dashed lines). There is good agreement between the predicted mean (a) and the prediction of Sibyll 2.3c, however there is a clear energy dependent bias in $\sigma\left(X_{\max}\right)$ (b) compared with the prediction of Sibyll 2.3c.

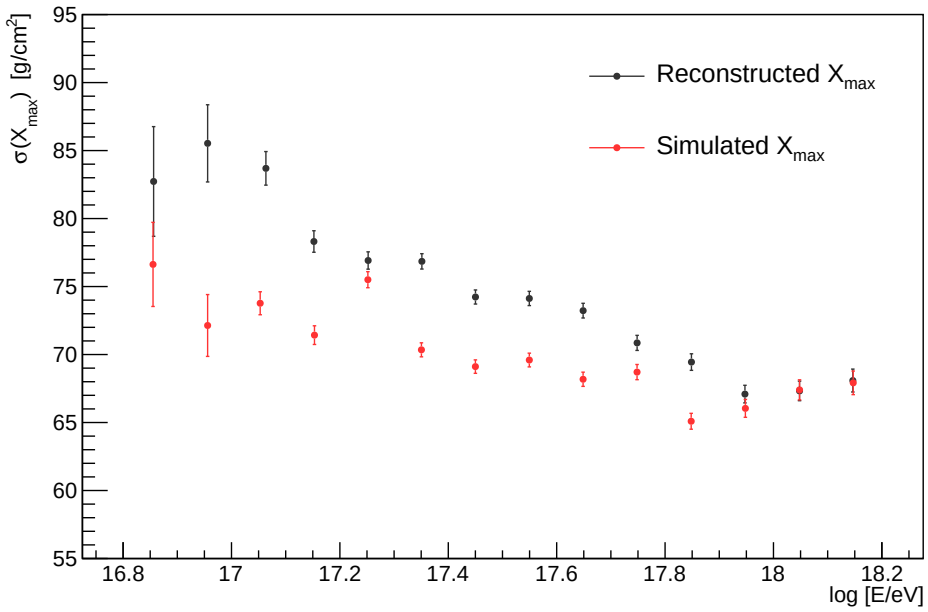


Figure 5.14. A comparison between $\sigma\left(X_{\mathrm{max}}\right)$ as calculated using the true (simulated) X_{max} distribution (red) and the reconstructed X_{max} distribution. See the discussion in the text for details.

through the reconstruction, this suggests that the effect is due to an additional unknown reconstruction bias which causes the width of the X_{max} distribution to be narrower than expected. Upon further investigation, the bias was found to be caused by an X_{max} dependent reconstruction bias. The average X_{max} reconstruction bias as a function of true X_{max} is shown in Figure 5.15. Importantly, the negative slope of the bias means that deep X_{max} values are, on average, reconstructed shallower, and shallow X_{max} values are reconstructed deeper. Therefore, the result of the bias is to narrow the reconstructed X_{max} distribution, thereby decreasing the observed $\sigma\left(X_{\mathrm{max}}\right)$: exactly the observation in Figures 5.12b and 5.13b. If instead the slope of the reconstruction bias were positive, then the effect would be the increase the width of the X_{max} distribution. This X_{max} dependent reconstruction bias can be explained by considering events with a true (simulated) X_{max} close to, or just outside the FD FoV. These events tend to be incorrectly reconstructed inside the FoV. This also explains the dependence on $\sigma\left(X_{\mathrm{max}}\right)$; a pure iron X_{max} distribution is narrower and therefore the bulk of the distribution is unaffected. In contrast, a pure proton X_{max} distribution is wider and events are therefore more likely to fall near or outside the detector FoV, and hence more likely to be affected by the X_{max} bias.

It is not possible to correct X_{max} for this bias on an event-by-event basis as for the reconstruction bias as a function of energy in Section 5.6. This is because such a correction would require knowledge of the true X_{max}; an impossibility for the measured data. These biased events are not removed by the selection cuts

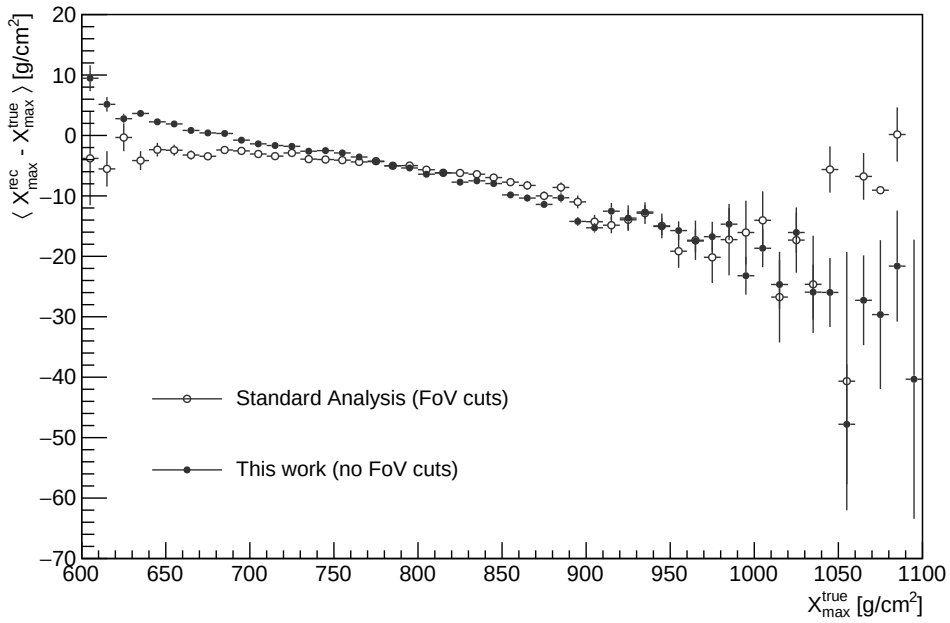


Figure 5.15. Average $X_{\max}$ bias as a function of true $X_{\max}$ for the standard FD data set. Shown as solid markers is the bias for the MC events used in this analysis without FoV cuts and shown as hollow markers is the bias for the MC events used in the standard $X_{\max}$ analysis with FoV cuts. The bias is similar in both cases, suggesting that the relaxed selection cuts, and lack of FoV cuts in this work does not cause the bias.

described in Section 5.5 since the reconstruction does not fail and the reconstruction uncertainties are small enough to pass the cut thresholds. One solution would be to introduce stricter cut thresholds on the data, in order to maintain a quality, unbiased event sample. However, to remove this bias would result in the loss of a significant number of events. The goal of this analysis is to increase statistics and extend the energy range of the $X_{\max}$ analysis, therefore the approach taken in this work will be to instead, correct the $\sigma\left(X_{\max}\right)$ bias. Furthermore, following the discovery of this bias, the standard analysis was checked for the same bias. It is shown in Figure 5.15, that even applying the (stricter) standard cuts, including the FoV cuts, does not remove the bias. Therefore, even the standard analysis requires such a correction. An end-to-end correction for the $\sigma\left(X_{\max}\right)$ bias is discussed in Section 5.7.2.

5.7.2 End-to-end σ Bias Correction

As discussed in Section 5.7.1, a bias exists in the calculated $\sigma\left(X_{\max}\right)$ compared to the theoretical $\sigma\left(X_{\max}\right)$ from the simulated hadronic model, Sibyll 2.3c. The bias is due to an $X_{\max}$ reconstruction bias which depends on the true $X_{\max}$. Since a correction on an event-by-event basis is impossible, an end-to-end correction will be derived in this section and tested using the same MC data sets used in Section 5.7.1.

Rather than determine an end-to-end correction for each composition individually, a correction as a function of the reconstructed $\sigma\left(X_{\text{max}}\right)$ will be derived. The benefit of this, is that the correction will not inherently depend on the composition of the data, only on $\sigma\left(X_{\text{max}}\right)$. This reduces the reliance on the fitted fractions and avoids the requirement of an iterative process to determine the final composition fractions of the data. To this end, the difference between the calculated $\sigma\left(X_{\text{max}}\right)$ and the prediction for Sibyll 2.3c is calculated for each of the simulated compositions; proton, helium, nitrogen, and iron. Additionally, the residuals for each combination of 50% composition mixes are calculated. The idea of producing composition mixes, is to create a large data set which covers a wide range of possible values of $\sigma\left(X_{\text{max}}\right)$. This provides a good constraint on the bias as a function of $\sigma\left(X_{\text{max}}\right)$ and $\log E$. In each energy bin, the residual between the reconstructed $\sigma\left(X_{\text{max}}\right)$, denoted σ_{rec} and the theoretical $\sigma\left(X_{\text{max}}\right)$, denoted σ_{model} is calculated for each composition mix. An example of this is shown in Figure 5.16 where the residual is plotted as a function of σ_{rec} for the energy range $19.1 < \log E/\text{eV} < 19.2$. This procedure is repeated for each energy bin. In order to reduce the effects of fluctuations from one energy bin to another, the correction is also parametrised as a function of $\log E$. This bias correction has the form

$$\Delta\left[\sigma\left(X_{\text{max}}\right)\right] = \begin{pmatrix} 1 \\ X_{\text{max}} \end{pmatrix}^{T} \cdot \left[\begin{pmatrix} a & b \\ d & d \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \log E \end{pmatrix} \right]. \tag{5.48}$$

The value of the correction as a function of σ_{rec} and $\log E$ is shown in Figure 5.17. The correction is largest at higher energy, as expected. Note than the correction is small for pure iron composition compared to pure proton for example due to its narrow width. Since the width of the pure iron distribution is much narrower, the effect of the X_{max} reconstruction bias is less. This is because the reconstruction bias is relatively constant over the narrow range of the iron distribution. That is, the X_{max} distribution is contained within a smaller range of Figure 5.15. The correction becomes as large as 8 g/cm^2 for high energy and heavier compositions. It is worth considering that the expected $\sigma\left(X_{\text{max}}\right)$ of the measured data from previous analysis (see Section 2.5) is larger at low energy ($\sim$60 g/cm^2 below $10^{18.5}$ eV), and progressively smaller at higher energy. The measured $\sigma\left(X_{\text{max}}\right)$ reaches $\sim$25-30 g/cm^2 at $10^{19.5}$ eV where the bias correction is $\sim$3 g/cm^2. Therefore, it is expected that the actual bias correction which will be applied to the measured $\sigma\left(X_{\text{max}}\right)$ is small. This is discussed further in Section 5.9.

The derived correction has been tested by applying it to the biased $\sigma\left(X_{\text{max}}\right)$ calculated for the MC events used in Section 5.7.1. Shown in Figures 5.18 and 5.19 is the fully corrected $\sigma\left(X_{\text{max}}\right)$ for the HeCo and standard FD data sets, respectively. Good agreement can be seen between the calculated $\sigma\left(X_{\text{max}}\right)$ and the prediction of Sibyll 2.3c for both data sets, and for all composition scenarios.

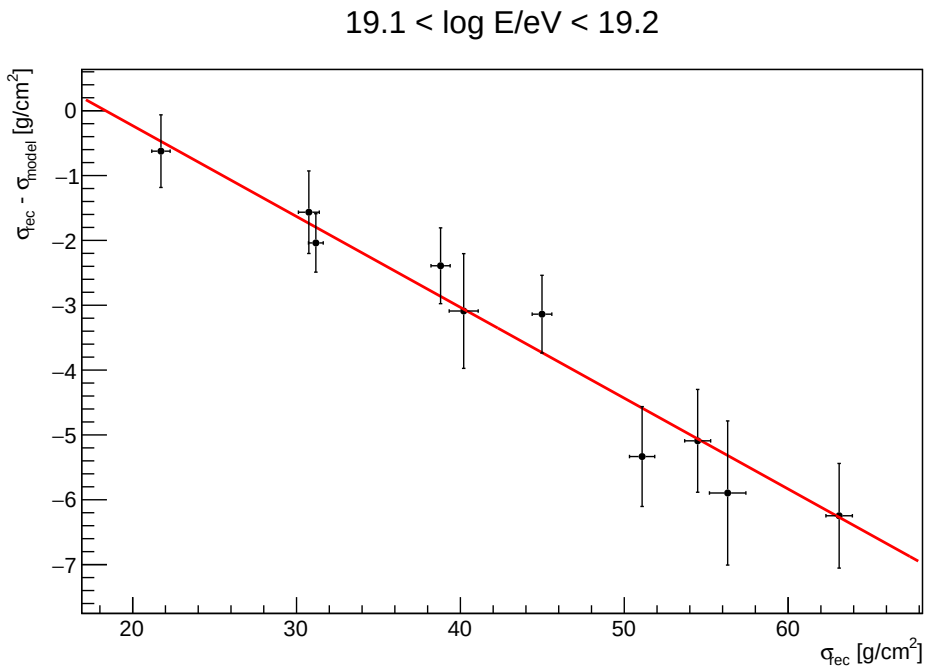


Figure 5.16. An example of the residual between the reconstructed $\sigma\left(X_{\max}\right)$ and the model prediction from Sibyll 2.3c as a function of $\sigma\left(X_{\max}\right)$ for the energy range $19.2 < \log E < 19.2$. This is repeated in all energy bins.

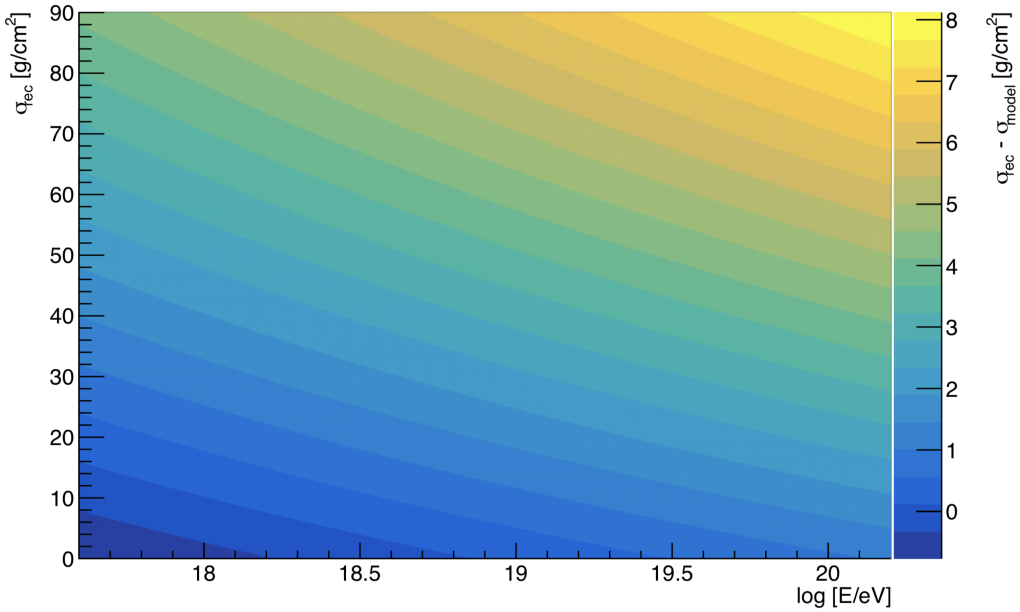


Figure 5.17. The value of the $\sigma\left(X_{\max}\right)$ bias correction as a function of reconstructed $\sigma\left(X_{\max}\right)$ and energy.

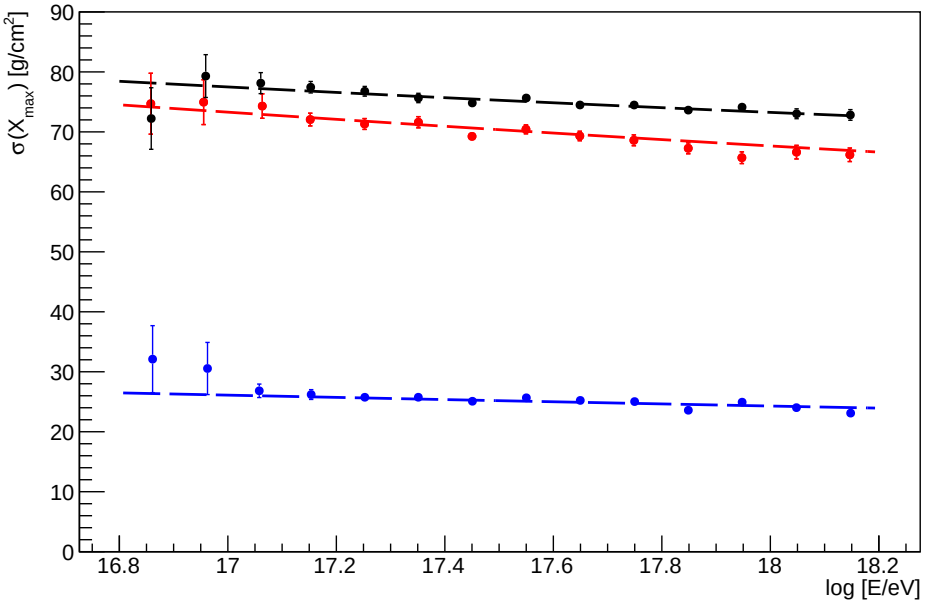


Figure 5.18. Standard deviation of the MC X_{max} distributions from the HeCo data set for pure proton (red), pure iron (blue) and a mix of 50% each of proton and iron (black) with the σ bias correction applied. The standard deviation is compared to the prediction of the simulated hadronic model, Sibyll 2.3c (dashed lines). The previous energy dependent bias in $\sigma\left(X_{\mathrm{max}}\right)$ compared with the prediction of Sibyll 2.3c is no longer present, indicating that the σ bias correction is valid.

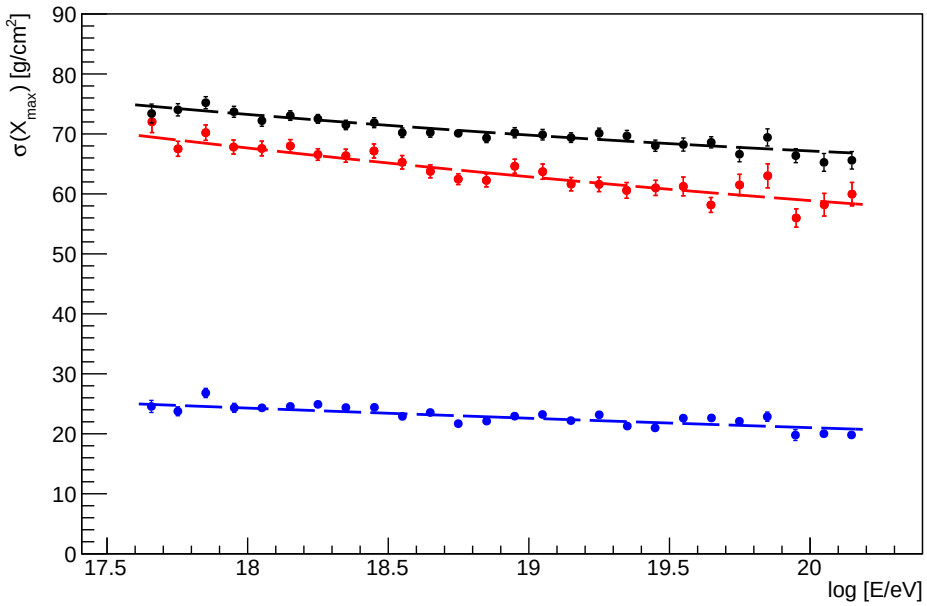


Figure 5.19. Standard deviation of the MC $X_{\max}$ distributions from the standard FD data set for pure proton (red), pure iron (blue) and a mix of 50% each of proton and iron (black) with the σ bias correction applied. The standard deviation is compared to the prediction of the simulated hadronic model, Sibyll 2.3c (dashed lines). The previous energy dependent bias in $\sigma\,(X_{\max})$ compared with the prediction of Sibyll 2.3c is no longer present, indicating that the σ bias correction is valid.

5.8 Further Resolution Corrections and Systematics Uncertainties

To analyse the measured $X_{\max}$ distributions as a function of energy and compare to predictions of hadronic interaction models, it is important to understand the effect of the corrections, systematic uncertainties and additional resolution effects from the detector, analysis method and the atmosphere which are not present in the MC simulations. The systematic uncertainties and resolution corrections of the HeCo and standard FD data sets are calculated independently. Unless otherwise specified, each component of the systematic uncertainties are added in quadrature to obtain the total systematic uncertainty.

5.8.1 Profile Reconstruction Systematic Uncertainties

Each of the systematic uncertainties associated with the Gaisser-Hillas profile reconstruction are discussed in detail below, and a summary of each uncertainty is shown in Figure 5.20.

5.8.1.1 $X_{\max}$ Reconstruction Bias

The $X_{\max}$ reconstruction bias is determined in Section 5.6.2 using the composition fractions calculated in Section 5.6.1. The $X_{\max}$ reconstruction bias is fully corrected using Equation (5.43). It is expected that the reconstruction bias observed in simulated events would also affect the data in the same way. However, it may well be that the bias is not due to shortcomings of the reconstruction, but rather an incompatibility of the simulation and reconstruction. Therefore, a conservative approach is taken, and a one-sided systematic uncertainty is assigned

$$\delta\left(X_{\max}\right)_{\text{rec bias}} = +\langle \Delta X_{\max} \rangle. \tag{5.49}$$

5.8.1.2 Gaisser-Hillas Fit Function

In addition to the $X_{\max}$ reconstruction bias correction, there are systematic uncertainties associated with the choice of the Gaisser-Hillas profile for the reconstruction procedure and constraints of the shape parameters, X_0 and λ (now R and L: See Section 3.5.2.7). Previous studies have shown that the choice of the Gaisser-Hillas function for the profile reconstruction instead of an alternative form based on a Gaussian in shower age shifts $X_{\max}$ on average by $\leq$4 g/cm^2 [199]. The constraint

of the Gaisser-Hillas profile shape parameters results in a systematic uncertainty in X_{max} of [196]

$$\Delta X_{\text{constr}} \leq [12.1 - 0.47 \lg(E/\text{eV})] \text{ g/cm}^2. \tag{5.50}$$

Since the systematic uncertainty due to the constraints of the shape parameters, and the systematic uncertainty due to the choice of the Gaisser-Hillas function are both settings in the profile reconstruction, it is possible they are correlated. Therefore, only the maximum of these effects will contribute to the systematic uncertainty of X_{max}:

$$\delta\left(X_{\text{max}}\right)_{\text{rec}} = \max\left[\delta\left(X_{\text{max}}\right)_{\text{GH}}, \delta\left(X_{\text{max}}\right)_{\text{constr.}}\right] \tag{5.51}$$

where

$$\delta\left(X_{\text{max}}\right)_{\text{GH}} = \pm 4 \text{ g/cm}^2 \tag{5.52}$$

and

$$\delta\left(X_{\text{max}}\right)_{\text{constr.}} = \pm[12.1 - 0.47 \lg(E/\text{eV})] \text{ g/cm}^2. \tag{5.53}$$

While the recent change of the Gaisser-Hillas shape parameters X_0 and λ to R and L is not expected to significantly change the magnitude of the systematic uncertainties associated with the profile fit, future work should consider the possibility by studying the systematic effects of R and L constraints for high quality events which can be reconstructed in full, with R and L freely fit.

5.8.1.3 Lateral Width Correction

The lateral width bias arises from the reconstruction procedure used to calculate the longitudinal shower profile. The lateral width correction is a parametrisation of the light outside the shower track across the FD camera [199, 200]. As discussed in Section 3.5.2.5, only pixels within an angle ζ of the SDP track are considered in the calculation of the total photon flux at the aperture. The algorithm only selects pixels close to the SDP because more distant pixels are dominated by noise, and their inclusion would reduce the overall signal-to-noise ratio. The lateral width correction is applied to the data during the reconstruction procedure to account for this bias. The difference between X_{max} with and without the lateral width correction applied during the reconstruction procedure is

$$\Delta_{\text{LWcorr}} = 14.8 \left[\exp\left(\frac{\lg(E/\text{eV}) - 18.68}{0.43}\right) + 1\right]^{-1} \text{ g/cm}^2. \tag{5.54}$$

The lateral width correction was shown to have a bias when applied to simulated showers [199]. In this analysis, X_{max} will be corrected for the bias by

$$X_{\text{max}}^{\text{true}} = X_{\text{max}}^{\text{rec}} - b_{\text{LWcorr}} \tag{5.55}$$

where

$$b_{\mathrm{LWcorr}} = 6.5 \left[\exp\left(\frac{\lg(E/\mathrm{eV}) - 18.23}{0.41} \right) + 1 \right]^{-1} \mathrm{g/cm^2}. \tag{5.56}$$

This bias correction as well as the full remaining difference in Equation (5.54) will be assigned as a systematic uncertainty:

$$\delta\left(X_{\mathrm{max}}\right)_{\mathrm{LWcorr}} = {}^{+b_{\mathrm{LWcorr}}}_{-(\Delta_{\mathrm{LWcorr}} - b_{\mathrm{LWcorr}})} \,. \tag{5.57}$$

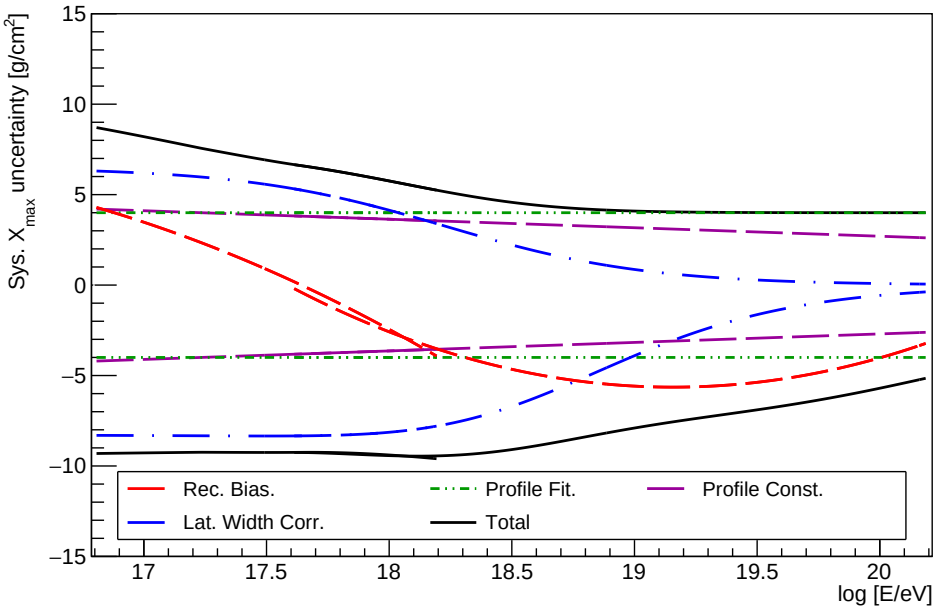


Figure 5.20. The X_{max} systematic uncertainties associated with the Gaisser-Hillas profile reconstruction for the HeCo data set ($16.8 \leq \log E \leq 18.2$) and the standard FD data set ($17.6 \leq \log E \leq 20.2$). Each component of the profile systematic uncertainty is shown; the X_{max} reconstruction bias (red), the choice of the Gaisser-Hillas function (green), the Gaisser-Hillas shape parameter constraints (magenta), and the lateral width correction (blue). Note the good agreement between the HeCo data set and the standard data set in the overlapping energy region ($17.6 \leq \log E \leq 18.2$). The total systematic uncertainty is shown in black. See the text for details of each uncertainty.

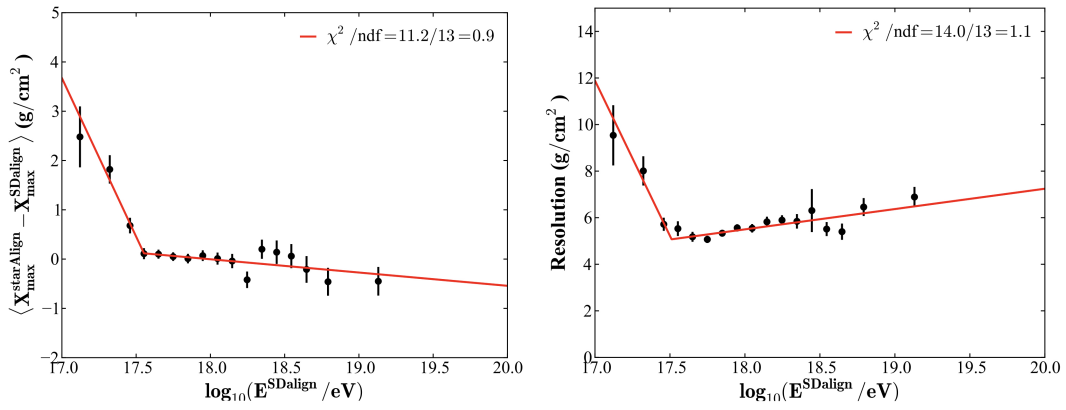
5.8.2 Detector Uncertainties

Uncertainties associated with the detectors are important to the X_{max} reconstruction. Telescope alignment, calibration and time synchronisation can all affect the X_{max} reconstruction. Therefore, the uncertainties in such measurements must be taken into account in the overall systematic uncertainty of X_{max}. The detector uncertainties which contribute to the X_{max} systematic uncertainties are discussed below.

5.8.2.1 Telescope Alignment

The misalignment of the FD telescopes can lead to a difference in the reconstructed $X_{\max}$. For example, if an event is reconstructed by an FD telescope which is misaligned in elevation, then $X_{\max}$ would appear to be deeper or shallower than it really was depending on the direction of the misalignment. Therefore, it is important to obtain a precise measurement of the alignment.

Several different methods are used to measure the alignment of the FD telescopes. The alignment can be measured based on a cross-calibration with high quality hybrid events measured by both the FD and SD [201]. Other methods of measuring the alignment include tracking the movement of stars across the FD camera [202, 203] and laser shots from the CLF or XLF [204, 205]. Shown in Figures 5.21a and 5.21b is the effect on the measured $X_{\max}$ and the contribution to the $X_{\max}$ resolution when changing from the alignment measurements with star tracks to alignment with hybrid events. High energy events which are, on average, further from the detector are affected more by the misalignment, while closer, low energy events are less affected. However, at the lowest energies the data are dominated by HeCo events. The misalignment between HEAT and Coihueco can adversely affect the SDP reconstruction, especially at the lowest energies where the track length in the FD cameras is shorter. Hence the systematic uncertainty increases for low energy showers.



(a) Systematic uncertainty associated with different alignment measurements.

(b) Contribution to the $X_{\max}$ resolution by different alignment measurements.

Figure 5.21. Difference in $X_{\max}$ between different telescope alignment measurements; alignment with SD measurements (SDalign) and star alignment (starAlign) [193].

The difference in X_{max} between the SD align and star align measurements can be parametrised as a function of energy, and is given by

$$\Delta\left(X_{\text{max}}\right)_{\text{align}} = \begin{cases} 0.12 - 6.43\left(\log E - 17.55\right)\ \text{g/cm}^2 & \log E < 17.55 \\ 0.12 - 0.27\left(\log E - 17.55\right)\ \text{g/cm}^2 & \text{otherwise} \end{cases} \tag{5.58}$$

and the contribution to the X_{max} resolution can be parametrised by

$$\sigma\left(X_{\text{max}}\right)_{\text{align}} = \begin{cases} 5.07 - 13.34\left(\log E - 17.51\right)\ \text{g/cm}^2 & \log E < 17.51 \\ 5.07 - 0.87\left(\log E - 17.51\right)\ \text{g/cm}^2 & \text{otherwise} \end{cases} \tag{5.59}$$

It is unclear which of the two alignment measurements is most accurate, therefore half of the difference in Equation (5.58) is corrected and the other half is assigned to a systematic uncertainty. Therefore, the contribution to the total X_{max} systematic uncertainty is given by

$$\delta\left(X_{\text{max}}\right)_{\text{align}} = \pm\frac{1}{2}|\Delta\left(X_{\text{max}}\right)_{\text{align}}|, \tag{5.60}$$

and the contribution to the X_{max} resolution is given by

$$\sigma\left(X_{\text{max}}\right)_{\text{align}} = \frac{1}{2}\sigma\left(X_{\text{max}}\right)_{\text{align}} \pm \frac{1}{2}\sigma\left(X_{\text{max}}\right)_{\text{align}}. \tag{5.61}$$

5.8.2.2 Pixel Calibration

An uncertainty in the absolute calibration of the FD affects the energy scale, while the X_{max} scale in unaffected. However, an elevation dependence in the Cal A (Section 3.2.2.3) relative calibration can affect X_{max}. Previous studies [196, 199] have shown this to have a small effect on X_{max} of ≤ 1 g/cm^2. Therefore, the systematic uncertainty in X_{max} associated with the pixel calibration is taken as

$$\delta\left(X_{\text{max}}\right)_{\text{cal}} = \pm 1\ \text{g/cm}^2. \tag{5.62}$$

5.8.2.3 SD-FD Time Offset

The time synchronisation between the SD and the FD is important for accurate hybrid geometry reconstruction. The uncertainty of the reconstructed shower axis within the SDP is less than 0.6°. However, if the FD and SD clocks were not synchronised, then the shower geometry would be systematically biased and hence cause a systematic bias in the X_{max} reconstruction. In order to maintain an angular resolution of 0.6°, any time offset between the FD and SD is required to be less than 100 ns. The offset can be measured with a precision better than 50 ns using inclined lasers and

golden hybrid events [196]. Selection cuts can reduce the effect of the time offset. Imposing a minimum viewing angle on the shower, for example, rejects events with a shower axis pointing towards the FD. Such showers are very short-duration and the time offset between the FD and SD becomes more significant. Events with a viewing angle less than 20° are rejected. This also has the effect of rejecting events with a large Cherenkov light fraction. The systematic effect of a time offset was studied by reconstructing events with an artificial time offset between the FD and SD. A time offset of 100 ns between the FD and the SD results in a systematic shift in X_{max} by up to 3.5 g/cm^2 for the HeCo energy range, and up to 2 g/cm^2 for the standard FD energy range [196, 199]. A conservative approach is taken by assigning the maximum shift as a systematic uncertainty in X_{max}. For the HeCo data set this becomes

$$\delta\left(X_{\mathrm{max}}\right)_{\delta t}^{\mathrm{HeCo}} = \pm 3.5 \text{ g/cm}^2 \tag{5.63}$$

and for the standard FD data set

$$\delta\left(X_{\mathrm{max}}\right)_{\delta t}^{\mathrm{FD}} = \pm 2 \text{ g/cm}^2. \tag{5.64}$$

5.8.2.4 HeCo Energy Calibration

The relative calibration between HEAT and Coihueco is extremely important for the accurate reconstruction of events which cross the FoV of both telescopes. Currently the energy calibration is determined using a series of data-driven studies which compares measurements between HEAT and Coihueco. In the simplest case, the energies of showers are measured independently by HEAT and Coihueco during time periods where the HEAT telescopes are operated in their downwards mode. These events are used to estimate the relative calibration based on their relative energies [206]. The energy ratio of such events changes with time as shown in Figure 5.22.

In addition to the cross-calibration using data where HEAT operated in downwards mode, events can be reconstructed with an additional parameter in the profile fit which controls the relative calibration between HEAT and Coihueco. The results of this study are shown in Figure 5.23 for data where HEAT was in both upwards and downwards mode. These results are compared to the energy ratio of events from Figure 5.22. There are significant differences between the results of the three data sets in Figure 5.23. The origin of the bias is still under investigation. A correction factor is applied to the results of the modified profile fits in order to shift the results from the profile fits to agree with the direct measurements of HEAT in downwards mode. The scaled results are compared in Figure 5.24 and show very good agreement. The data are then fit with a second-order polynomial in each

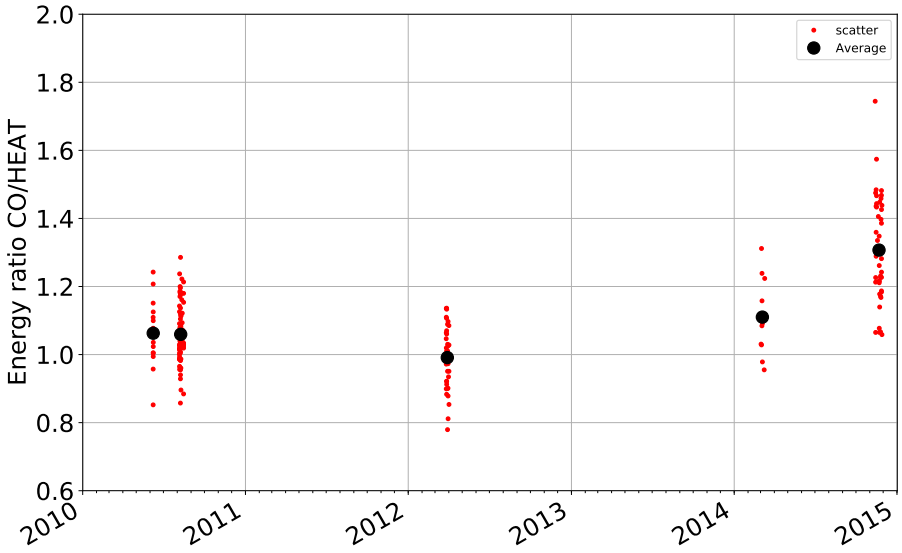


Figure 5.22. The relative energy calibration of HEAT and Coihueco as determined by independent measurements of the shower energy from time periods where HEAT operated in downwards mode [206].

data epoch. The result of this fit is also shown in Figure 5.24. The fitted time dependent relative calibration is applied to the data during reconstruction to correct for relative calibration differences over time. It is therefore expected that any X_{max} bias associated with the relative calibration has been corrected for. However, a residual spread in the X_{max} distribution is present following the correction. This is possibly a consequence of not all telescopes conforming to the average. The residual spread in the X_{max} distribution must be subtracted in quadrature from $\sigma\left(X_{\text{max}}\right)$ akin to the detector resolution. This additional resolution correction is given by

$$\sigma\left(X_{\text{max}}\right)_{\text{HeCo calib.}} = \begin{cases} 17.75 + 25\left(\log E - 17.8\right)^2 \text{ g/cm}^2 & \log E < 17.7 \\ 18 \text{ g/cm}^2 & \text{otherwise} \end{cases} \tag{5.65}$$

The HeCo energy calibration resolution correction is shown in comparison to the detector resolution and atmospheric resolution effects in Figure 5.35. The HeCo relative calibration is a topic of ongoing work in Auger. A more sophisticated technique is currently under development to determine the relative calibration which will account for the additional differences on a telescope-by-telescope basis.

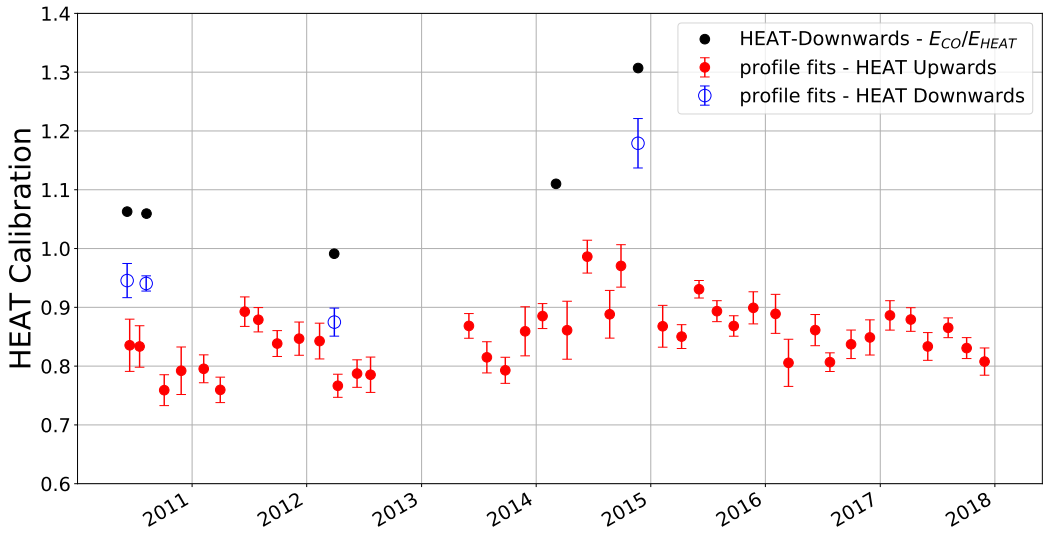


Figure 5.23. The relative energy calibration of HeCo as determined by a modified profile fit during the reconstruction compared with the direct measurements with HEAT in downwards mode from Figure 5.22 (black). The relative calibration is fit separately for events with HEAT in upwards mode (red) and downwards mode (blue). Note the offsets between each data set [206].

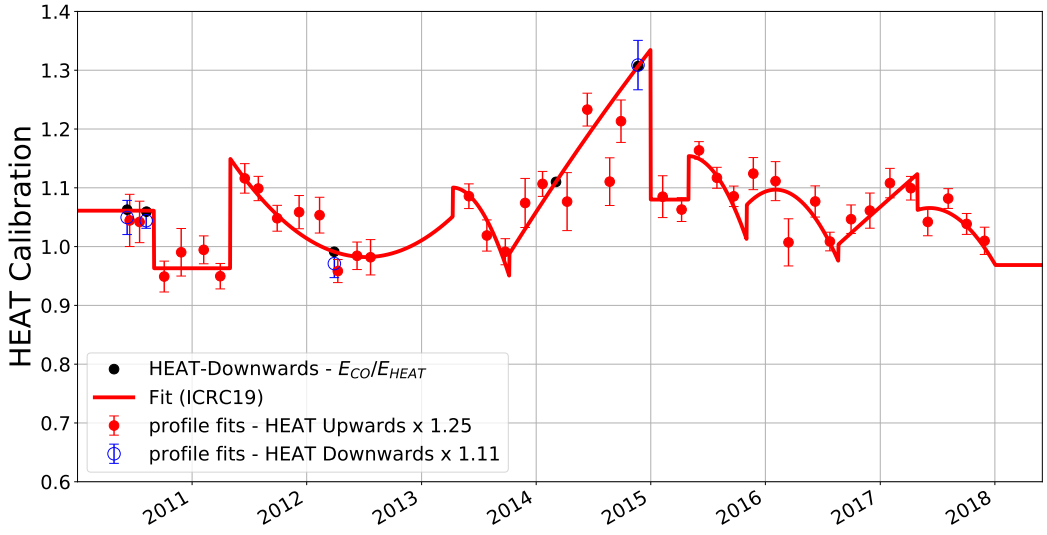


Figure 5.24. The corrected HeCo relative energy calibration as a function of time with each data epoch fitted with a second-order polynomial (red line) [206].

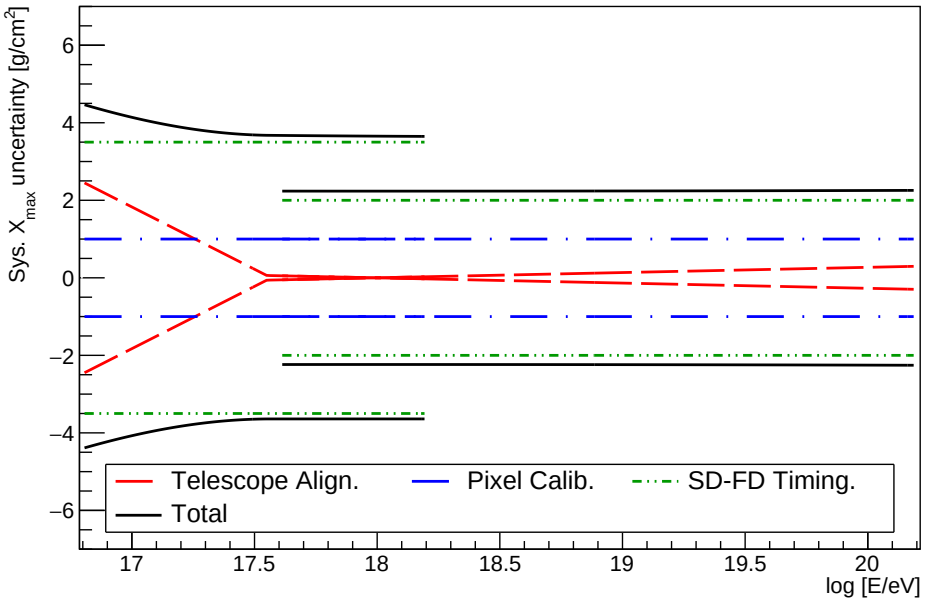


Figure 5.25. The $X_{\max}$ systematic uncertainties associated with the detector. The systematic uncertainties are shown for the HeCo data set ($16.8 \leq \log E \leq 18.2$) and the standard FD data set ($17.6 \leq \log E \leq 20.2$). Each component of the detector systematic uncertainties are shown; the telescope alignment (red), pixel calibration (blue), FD-SD time synchronisation (green). The total systematic uncertainty is shown in black. See the text for details of each uncertainty.

5.8.3 End-to-end Correction Uncertainty

As discussed in Section 5.7.1, a bias exists in the calculated $\sigma\left(X_{\text{max}}\right)$ compared to the theoretical prediction from the simulated hadronic model, Sibyll 2.3c. An end-to-end correction was parametrised as a function of the reconstructed $\sigma\left(X_{\text{max}}\right)$ in Section 5.7.2. This means that the correction is independent of composition and can therefore be applied to the data. The correction is applied in full, since it is expected that the X_{max} bias would affect the data in the same way as it affects the MC simulated events. However, it may well be that the bias is not due to shortcomings of the reconstruction, but rather an incompatibility of the simulation and reconstruction. A conservation approach is taken, and a one-sided systematic uncertainty is assigned to account for the possibility that the bias does not exists in the data. The systematic uncertainty assigned to $\sigma\left(X_{\text{max}}\right)$ is therefore

$$\delta\left[\sigma\left(X_{\text{max}}\right)\right]_{\text{end-to-end}} = ^{+0}_{-\Delta[\sigma(X_{\text{max}})]} \tag{5.66}$$

where $\Delta\left[\sigma\left(X_{\text{max}}\right)\right]$ is defined in Equation (5.48).

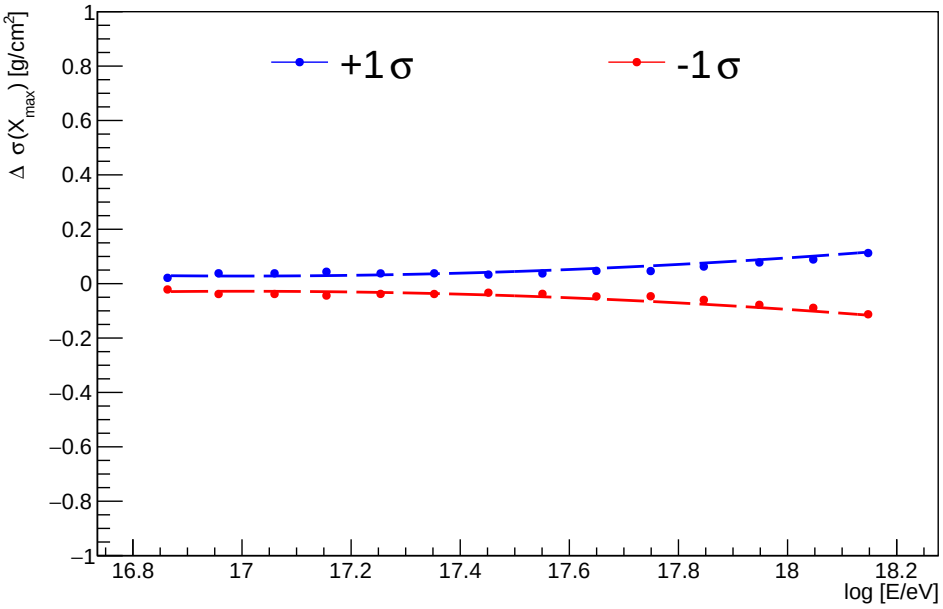
In addition to the one-sided systematic uncertainty based on the magnitude of the correction, the effect of the uncertainty in $\sigma\left(X_{\text{max}}\right)$ means that the value of the correction is also uncertain. To test the effect of the uncertainty in $\sigma\left(X_{\text{max}}\right)$ on the magnitude of the correction, the correction is evaluated at the $\pm 1\sigma$ values of $\sigma\left(X_{\text{max}}\right)$ and applied to the central value of $\sigma\left(X_{\text{max}}\right)$. The difference between $\sigma\left(X_{\text{max}}\right)$ with the correction based on the central value and the correction based on the $\pm 1\sigma$ values for the HeCo and standard FD data sets is shown in Figures 5.26a and 5.26b, respectively. The difference is negligible (< 0.2 g/cm^2), however the systematic bias is clearly energy dependent, so a systematic uncertainty has been parametrised as a function of energy. The results of these parametrisations are given by

$$\delta\left[\sigma\left(X_{\text{max}}\right)\right]^{\text{HeCo}}_{\text{end-to-end}} = \pm\left(19.34 - 2.27\log E + 0.067\log^2 E\right)\ \text{g/cm}^2 \tag{5.67}$$

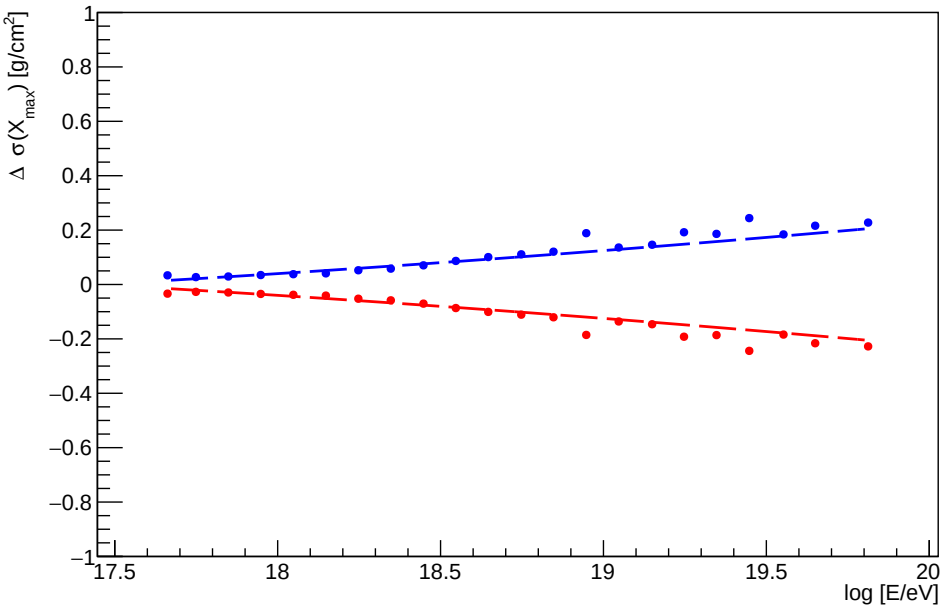
for the HeCo data set, and

$$\delta\left[\sigma\left(X_{\text{max}}\right)\right]^{\text{FD}}_{\text{end-to-end}} = \pm\left(1.24 - 0.21\log E + 0.008\log^2 E\right)\ \text{g/cm}^2 \tag{5.68}$$

for the standard FD data set. For completeness, these systematic uncertainties are added to the total.



(a)



(b)

Figure 5.26. The systematic uncertainty associated with the end-to-end bias correction on $\sigma\left(X_{\max}\right)$ for (a) the HeCo data set, and (b) the standard FD data set. The blue points correspond to the difference when applying the end-to-end correction based on the value of $\sigma\left(X_{\max}\right) + 1\sigma$, and the red points correspond to $\sigma\left(X_{\max}\right)$ - 1σ. The dashed lines represent the parametrisation of the difference as a function of energy. The equations are given in Equations (5.67) and (5.68).

5.8.4 X_{max} Acceptance Uncertainty

The systematic uncertainties associated with the kernel acceptance arise from several different sources. These include the uncertainties associated with the detector simulation, the kernel fit uncertainties and the settings of the kernel fit. Each of these are studied in detail below. A summary of the kernel acceptance systematic uncertainties is shown in Figure 5.31.

5.8.4.1 Energy Scale Uncertainty

In order to check the dependence of the detector simulation on the acceptance, the energy of the flat X_{max} MC events used to derive the X_{max} acceptance are shifted by the total energy scale uncertainty of $\pm 14\%$. This mimics the effect of the uncertainty of the light production and telescope efficiencies within the total energy scale uncertainty of the detector. The kernel X_{max} acceptance fits, and the entire analysis sequence is repeated with the energy-shifted acceptance. Finally, the difference between the moments which are calculated using the usual acceptance correction, and the energy shifted acceptance correction, are calculated. The results of this study is shown in Figure 5.27. The effects are small for both data sets; ~ 2 g/cm^2 for $\langle X_{\text{max}} \rangle$ and ~ 0.5 g/cm^2 for $\sigma\left(X_{\text{max}}\right)$. Therefore, the following systematic uncertainties are assigned to $\langle X_{\text{max}} \rangle$ and $\sigma\left(X_{\text{max}}\right)$, respectively:

$$\delta\left(X_{\text{max}}\right)_{\text{energy scale}} = \pm 2 \text{ g/cm}^2 \tag{5.69}$$

$$\delta\left[\sigma\left(X_{\text{max}}\right)\right]_{\text{energy scale}} = \pm 0.5 \text{ g/cm}^2. \tag{5.70}$$

5.8.4.2 Kernel Fit Uncertainties

The kernel fit does not automatically produce fit uncertainties like regression or likelihood fit functions. Instead the fit uncertainties must be obtained by propagating the statistical uncertainties from the X_{max} acceptance distribution through the kernel fit. In order to do this, the bootstrap procedure is used. This involves randomly sampling (with replacement) from the original X_{max} acceptance distribution to obtain a series of realisations of the distribution. The differences in the kernel function, and hence the acceptance correction are solely due to the statistical uncertainty in the acceptance distribution. In order to determine the systematic uncertainty in X_{max} associated with the fit uncertainty, the entire analysis procedure is repeated for each realisation of the X_{max} acceptance distribution and the X_{max} moments are subsequently calculated using each of the modified acceptance corrections. This produces a distribution of possible X_{max} moments, the variance of which is taken as the best estimate of the kernel fit uncertainty. The results of this study are shown in

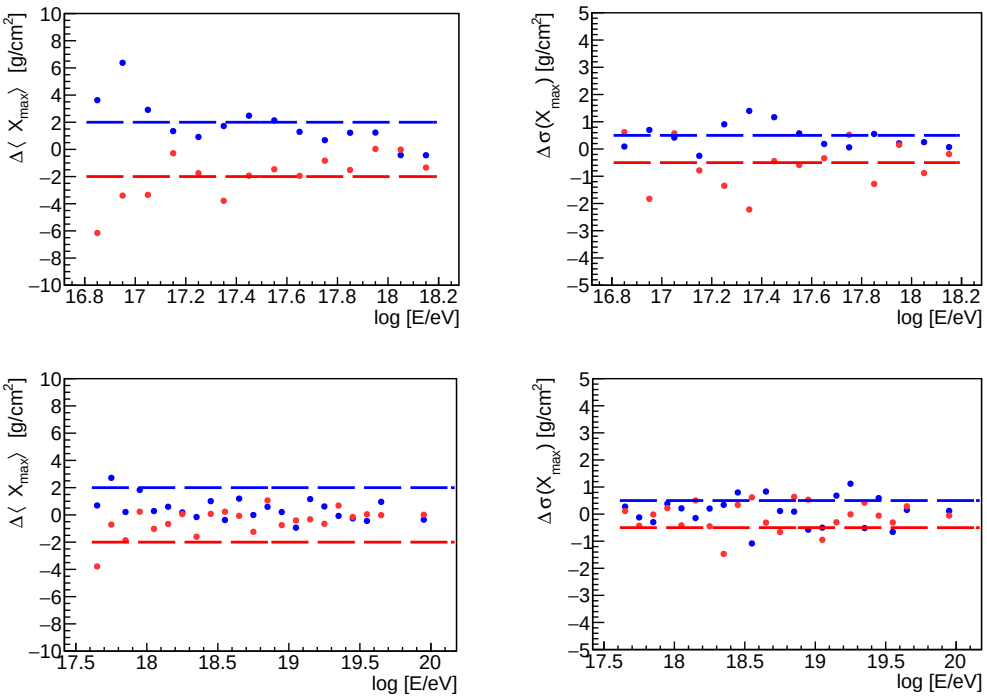


Figure 5.27. Changes of the $X_{\max}$ moments under a $\pm 14\%$ shift in the energy scale of the events used to determine the $X_{\max}$ acceptance correction. The blue points correspond to a shift of +14%, and the blue points to a shift of -14%. The upper row shows the change for the HeCo data set, and the lower row shows the change for the standard FD data set. The dashed lines correspond to the estimated systematic uncertainties assigned to the total for this effect.

Figure 5.28. While there is a clear energy dependence in the systematic uncertainty, its magnitude is small; < 2 g/cm^2 for both $\langle X_{\max} \rangle$ and $\sigma\left(X_{\max}\right)$ for both the HeCo and standard FD data sets. Nevertheless, the systematic uncertainty is parametrised as a function of energy for each case. The resulting parametrisations are given by

$$\delta\left(X_{\max}\right)^{\text{HeCo}}_{\text{fit uncert.}} = \pm \begin{cases} 0.433 + 0.1903\left(\log E - 17.17\right)\ \text{g/cm}^2 & \log E < 17.17 \\ 0.433 + 7.559\left(\log E - 17.17\right)\ \text{g/cm}^2 & \text{otherwise} \end{cases} \tag{5.71}$$

$$\delta\left[\sigma\left(X_{\max}\right)\right]^{\text{HeCo}}_{\text{fit uncert.}} = \pm \begin{cases} 0.282 + 0.155\left(\log E - 17.18\right)\ \text{g/cm}^2 & \log E < 17.18 \\ 0.282 + 3.701\left(\log E - 17.18\right)\ \text{g/cm}^2 & \text{otherwise} \end{cases} \tag{5.72}$$

$$\delta\left(X_{\max}\right)^{\text{FD}}_{\text{fit uncert.}} = \pm\left(5.118 - 0.242\log E\right)\ \text{g/cm}^2 \tag{5.73}$$

$$\delta\left[\sigma\left(X_{\max}\right)\right]^{\text{FD}}_{\text{fit uncert.}} = \pm\left(4.760 + 0.232\log E\right)\ \text{g/cm}^2. \tag{5.74}$$

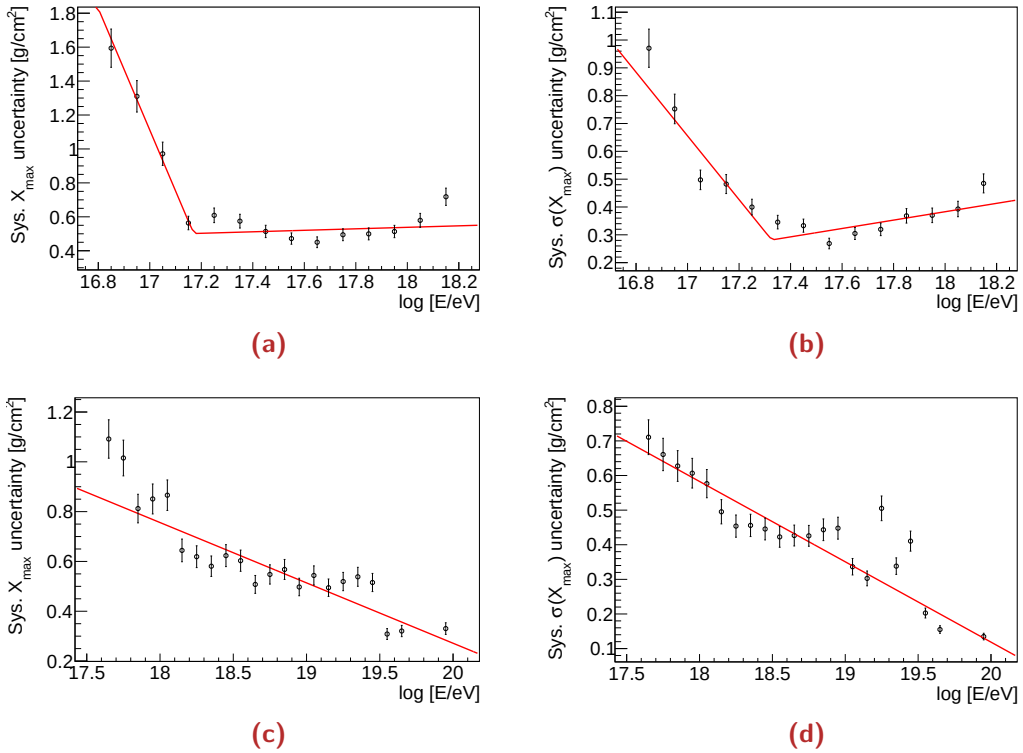


Figure 5.28. Kernel acceptance function fit uncertainties for (a) $\langle X_{\mathrm{max}} \rangle$ and (b) $\sigma\left(X_{\mathrm{max}}\right)$ for the HeCo data set, and (c) $\langle X_{\mathrm{max}} \rangle$ and (d) $\sigma\left(X_{\mathrm{max}}\right)$ for the standard FD data set. The uncertainties are determined using bootstrap sampling of the acceptance distributions. See text for details.

5.8.4.3 Kernel Smoothing

As described in Section 5.3, the kernel acceptance fit involves choosing a parameter, ρ to control the level of smoothing applied to the kernel fit. For this work $\rho = 1.5$ was chosen to balance the under and over-fitting in the bulk of the distribution and its tails in order to provide the best fit acceptance function while smoothing out statistical fluctuations in the X_{max} acceptance distribution. It is expected this will contribute some systematic uncertainty to the final X_{max} moments since the choice of ρ affects the acceptance function, and hence the acceptance correction applied to the X_{max} moments. To evaluate the magnitude of this systematic uncertainty, the value of ρ is systematically changed to four different trial values: 0.5, 1, 2 and 2.5. Outside of this range, the kernel function either significantly under or over-fits the acceptance distribution. Given these four options for ρ and the chosen value of $\rho = 1.5$, the entire analysis procedure was repeated using modified acceptance corrections derived for each of the values of ρ. The difference between the final X_{max} moments with $\rho = 1.5$ and $\rho = 0.5, 1, 2, 2.5$ are evaluated, the results of which

are shown in Figure 5.29. The differences are bracketed by the following functions for each of the X_{max} moments

$$\delta\left(X_{\mathrm{max}}\right)_{\mathrm{smooth}} = \pm \begin{cases} 121.7 - 7\log E \ \mathrm{g/cm^2} & \log E < 17.3 \\ 0.5 \ \mathrm{g/cm^2} & \text{otherwise} \end{cases} \tag{5.75}$$

$$\delta\left[\sigma\left(X_{\mathrm{max}}\right)\right]_{\mathrm{smooth}} = \pm \begin{cases} 52.4 - 3\log E \ \mathrm{g/cm^2} & \log E < 17.3 \\ 0.5 \ \mathrm{g/cm^2} & \text{otherwise.} \end{cases} \tag{5.76}$$

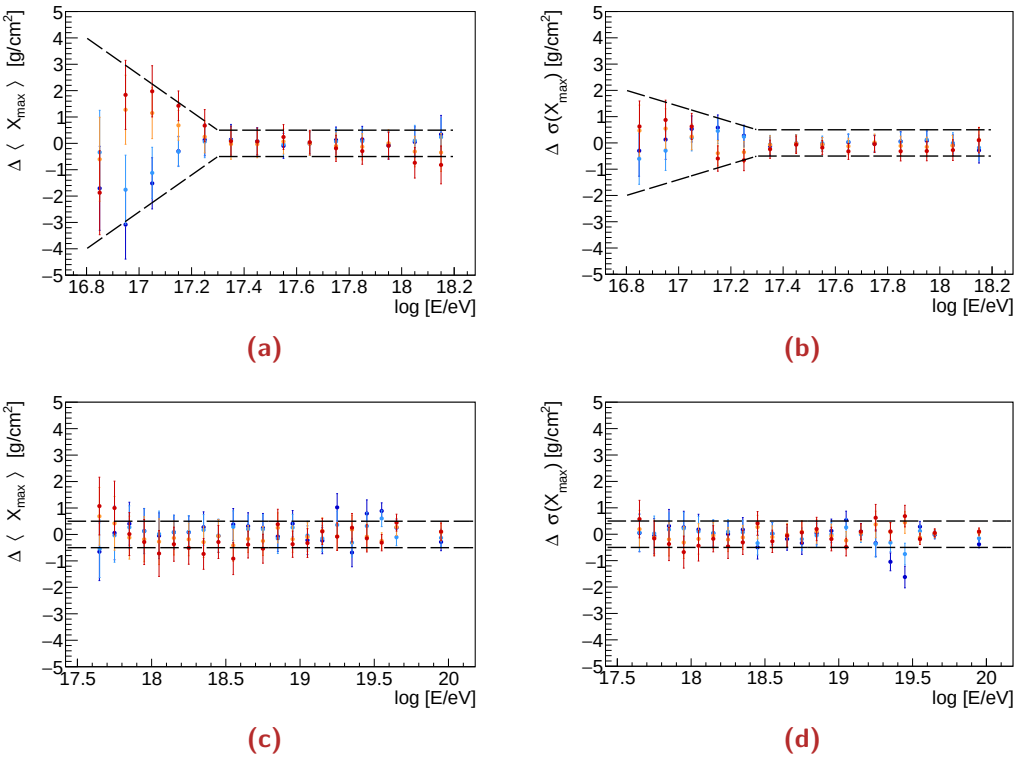


Figure 5.29. The systematic uncertainties associated with the kernel smoothing parameter, ρ, for (a) $\langle X_{\mathrm{max}} \rangle$ and (b) $\sigma\left(X_{\mathrm{max}}\right)$ for the HeCo data set, and for (c) $\langle X_{\mathrm{max}} \rangle$ and (d) $\sigma\left(X_{\mathrm{max}}\right)$ for the standard FD data set. The systematic uncertainties are determined by calculating the difference in the final X_{max} moments between $\rho = 1.5$ and $\rho = 0.5$ (dark blue), $\rho = 1$ (light blue), $\rho = 2$ (orange), and $\rho = 2.5$ (red). More extreme values of ρ are visually rejected due to under or over-fitting.

5.8.4.4 Acceptance Correction

While significant effort has been made to ensure the quality of the acceptance fits and any uncertainty is included as a systematic uncertainty, it is difficult to determine the uncertainty of the acceptance itself. Systematic uncertainties associated with the detector simulation remain unaccounted for. In the higher energy range where the

acceptance correction is small these effects are negligible, however at the lowest energies of this analysis the acceptance correction is extremely large. The X_{max} analysis is increasingly limited by the acceptance of the detector below $\sim 10^{17.5}$ eV. Therefore, a conservative systematic uncertainty of 50% of the acceptance correction will be taken into account in this analysis. The final X_{max} moments have been calculated with and without the acceptance correction applied in order to determine the magnitude of the correction as a function of energy. This is shown in Figure 5.30. The systematic uncertainty (50% of the correction) is very small (< 2 g/cm^2) above $10^{17.7}$ eV, however it increases up to $\sim$25 g/cm^2 at the lower energy limit of the analysis, $10^{16.8}$ eV. The systematic uncertainties are parametrised as a function of energy and given by

$$\delta\left(X_{\mathrm{max}}\right)^{\mathrm{HeCo}}_{\text{acc. corr.}} = \pm\frac{1}{2}\left(-8923.5 + 979.6\log E - 26.9\log^2 E\right)\ \mathrm{g/cm^2} \tag{5.77}$$

$$\delta\left[\sigma\left(X_{\mathrm{max}}\right)\right]^{\mathrm{HeCo}}_{\text{acc. corr.}} = \pm\frac{1}{2}\left(-2996.2 + 329.5\log E - 9.05\log^2 E\right)\ \mathrm{g/cm^2} \tag{5.78}$$

$$\delta\left(X_{\mathrm{max}}\right)^{\mathrm{FD}}_{\text{acc. corr.}} = \pm\frac{1}{2}\left(-1700.8 + 178.5\log E - 4.67\log^2 E\right)\ \mathrm{g/cm^2} \tag{5.79}$$

$$\delta\left[\sigma\left(X_{\mathrm{max}}\right)\right]^{\mathrm{FD}}_{\text{acc. corr.}} = \pm\frac{1}{2}\left(-419.8 + 46.6\log E - 1.28\log^2 E\right)\ \mathrm{g/cm^2} \tag{5.80}$$

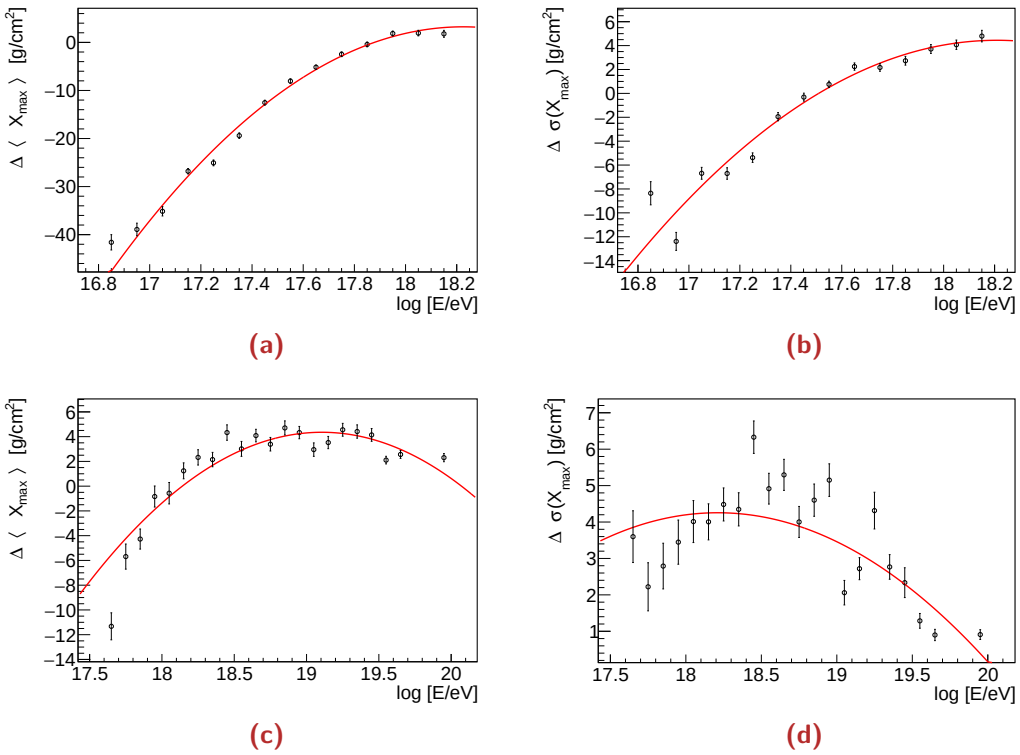
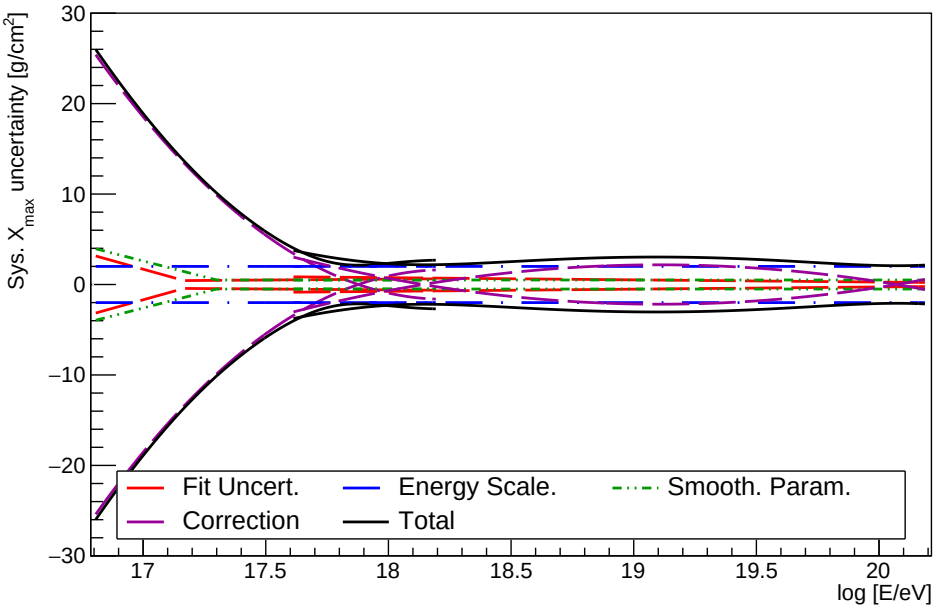
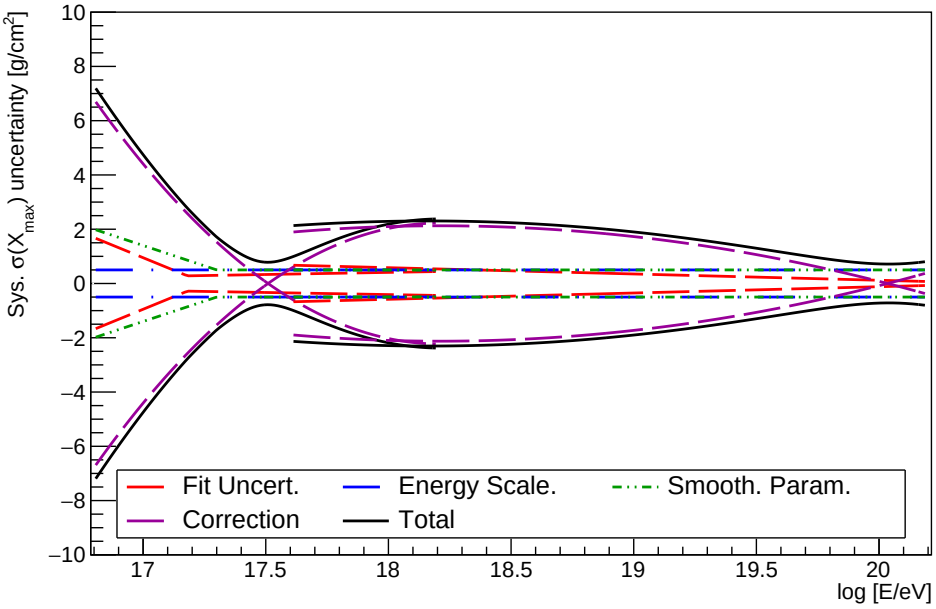


Figure 5.30. Changes of the final $X_{\max}$ moments with and without the acceptance correction applied for (a) $\langle X_{\max} \rangle$ and (b) $\sigma\,(X_{\max})$ of the HeCo dataset, and (c) $\langle X_{\max} \rangle$ and (d) $\sigma\,(X_{\max})$ of the standard FD dataset. 50% of this correction is conservatively assigned as a systematic uncertainty.



(a)



(b)

Figure 5.31. The systematic uncertainties in (a) $\langle X_{\mathrm{max}} \rangle$ and (b) $\sigma\left(X_{\mathrm{max}}\right)$ associated with the kernel acceptance correction. The systematic uncertainties are shown for the HeCo data set ($16.8 \leq \log E \leq 18.2$) and the standard FD data set ($17.6 \leq \log E \leq 20.2$). Each component of the systematic uncertainty is shown; the kernel fit uncertainties derived using bootstrap sampling (red), the energy scale (blue), smoothing parameter (green) and the conservative 50% uncertainty on the acceptance correction itself (magenta). The total systematic uncertainty is shown in black. See the text for details of each uncertainty.

5.8.5 Atmospheric Uncertainties and Contributions to Resolution

As discussed in Section 3.3, the atmosphere plays a pivotal role in FD measurements. Assumptions made about atmospheric conditions can significantly affect the reconstruction of events by the FD, and hence the reconstructed X_{max}. The atmospheric conditions are constantly monitored through measurements of the VAOD and are taken into account during reconstruction. However, uncertainty in the atmospheric measurements results in systematic uncertainty in X_{max}. The atmospheric effects also result in a broadening of the X_{max} distribution in addition to the usual shower-to-shower fluctuations. Therefore, these effects must be measured and used to correct the X_{max} distributions and their moments. Many studies of the effect of the atmosphere on the measured X_{max} distributions have been undertaken. Each contribution from the atmosphere to systematic uncertainties in X_{max} and X_{max} resolution are described below and summarised in Figures 5.32 and 5.33, respectively.

5.8.5.1 Molecular Atmosphere

Atmospheric conditions such as pressure, density and temperature have an impact on the molecular atmosphere. As discussed in Section 3.3, these atmospheric conditions are obtained from the Global Data Assimilation System (GDAS) and taken into account during the data reconstruction procedure. The influence of the molecular atmosphere on the X_{max} resolution can be estimated from the spread of the difference between shower reconstructions using GDAS and actual balloon soundings. According to [162, 199], the influence of the X_{max} reconstruction due to GDAS data on the X_{max} resolution is 2 g/cm^2 at 10^{18} eV and 3.5 g/cm^2 at 10^{20} eV. Assuming a linear change, it is deduced that the influence of the molecular atmosphere on the X_{max} resolution as a function of energy is

$$\sigma\left(X_{\text{max}}\right)_{\text{molAtmos}} = 2 + 0.75\left(\log E - 18\right)\ \text{g/cm}^2. \tag{5.81}$$

5.8.5.2 Fluorescence Yield

According to [192, 193, 199] the choice of fluorescence yield model introduces a systematic uncertainty on X_{max} of

$$\delta\left(X_{\text{max}}\right)_{\text{FY}} = \pm 0.4\ \text{g/cm}^2. \tag{5.82}$$

This is determined based on the change in X_{max} between the AIRFLY fluorescence yield model used by Auger, and other models including the Kaimoto, Keilhauer, and Nagano models [199].

5.8.5.3 Multiple Scattering

The systematic uncertainty of X_{max} originating from the multiple scattering correction is estimated by switching from the default Roberts multiple scattering model [207] to an alternative model from Pekala [208]. According to [209], switching models results in a change in $X_{\text{max}} \leq \pm 2$ g/cm^2. Therefore, the systematic uncertainty in X_{max} is given as

$$\delta\left(X_{\text{max}}\right)_{\text{ms}} = \pm 2 \text{ g/cm}^2. \tag{5.83}$$

In addition to the X_{max} shift, the variation of the typical aerosol size around the mean value used in the reconstruction adds a contribution to the detector resolution [210] of

$$\sigma\left(X_{\text{max}}\right)_{\text{ms}} = 1 \text{ g/cm}^2. \tag{5.84}$$

5.8.5.4 VAOD Statistical Uncertainty

The measurement of the VAOD by the CLF and XLF is averaged hourly, as described in Section 3.3. This average is used during shower reconstruction and the uncertainties are propagated into the calculation of the longitudinal shower profile. The uncertainty in the average VAOD therefore contributes to the X_{max} resolution. Higher energy showers, which are far from the detector suffer the most since changes in the aerosol content of the atmosphere are amplified due to aerosol scattering between the shower and the detector. For lower energy showers, which are closer to the detector aerosol scattering is negligible and the effects are smaller. According to [199], the average contribution of the VAOD statistical uncertainty on the X_{max} resolution is given by

$$\sigma\left(X_{\text{max}}\right)_{\text{VAOD, stat.}} = \sqrt{12\left[\exp\left(\frac{17.9 - \log E}{0.28} + 1\right)\right]^{-1}} \text{ g/cm}^2. \tag{5.85}$$

5.8.5.5 VAOD Systematic Uncertainty

The uncertainties associated with the CLF or XLF laser energy, FD calibration, and choice of reference night are correlated and contribute to the systematic uncertainty in the VAOD. To determine the effect of these uncertainties on X_{max} measurements, the data are reconstructed with the VAOD shifted by $\pm 1\sigma$. According to [199], the

systematic uncertainty of X_{max} due to the systematic uncertainty in the VAOD is given by

$$\delta\left(X_{\text{max}}\right)_{\text{VAOD, sys.}} = \pm 2\left[\exp\left(\frac{17.9 - \log E}{0.4} + 1\right)\right]^{-1} \text{ g/cm}^2. \tag{5.86}$$

The standard deviation of the X_{max} distribution due to the shifted VAOD is given by

$$\sigma\left(X_{\text{max}}\right)_{\text{VAOD, sys.}} = 2.7\left[\exp\left(\frac{17.4 - \log E}{0.6} + 1\right)\right]^{-1} \text{ g/cm}^2. \tag{5.87}$$

However, it is unclear if the standard deviation is simply due to statistical fluctuations of the X_{max} reconstruction due to fitting the shower profile with slightly different attenuation corrections [199]. Therefore, half of this contribution will be added to the detector resolution with an uncertainty accounting for the other half. That is

$$\sigma\left(X_{\text{max}}\right)_{\text{VAOD, sys.}} \Rightarrow \frac{1}{2}\sigma\left(X_{\text{max}}\right)_{\text{VAOD, sys.}} \pm \frac{1}{2}\sigma\left(X_{\text{max}}\right)_{\text{VAOD, sys.}} \tag{5.88}$$

5.8.5.6 VAOD Horizontal Uniformity

The VAOD measurements assume horizontal uniformity of atmospheric aerosol content. In order to determine the systematic uncertainty due to this simplification, events were re-reconstructed using the VAOD profile of different FD sites in [199]. The systematic and spread of the X_{max} reconstruction are given by

$$\delta\left(X_{\text{max}}\right)_{\text{VAOD, unif.}} = \pm 2.8 + 0.58\left(\log E - 18\right) \text{ g/cm}^2. \tag{5.89}$$

$$\sigma\left(X_{\text{max}}\right)_{\text{VAOD, unif.}} = \sqrt{\frac{\left(14\left[\exp\left(\frac{17.8-\log E}{0.65} + 1\right)\right]^{-1}\right)^2 - 2\sigma\left(X_{\text{max}}\right)^2_{\text{VAOD, stat.}}}{2}} \text{ g/cm}^2. \tag{5.90}$$

However, it is possible that the variance is overestimated since the aerosol data from a different FD site is less accurate than from the correct FD. The assumption of the contribution of any horizontal non-uniformity to the X_{max} resolution is somewhere between $\sigma\left(X_{\text{max}}\right)_{\text{VAOD, unif.}}/2$ and $\sigma\left(X_{\text{max}}\right)_{\text{VAOD, unif.}}$. Therefore, three-quarters of $\sigma\left(X_{\text{max}}\right)_{\text{VAOD, unif.}}$ is taken as the contribution to the X_{max} resolution, and one-quarter is assigned to the systematic uncertainty in $\sigma\left(X_{\text{max}}\right)$

$$\sigma\left(X_{\text{max}}\right)_{\text{VAOD, unif.}} \Rightarrow \frac{3}{4}\sigma\left(X_{\text{max}}\right)_{\text{VAOD, unif.}} \pm \frac{1}{4}\sigma\left(X_{\text{max}}\right)_{\text{VAOD, unif.}} \tag{5.91}$$

5.8.5.7 VAOD Normalisation

The aerosol content through the VAOD measurements are determined using the data normalised method, as discussed in Section 3.3. In the data normalised method, the profiles of the CLF and XLF laser shots are compared to reference night profiles which are measured on nights with presumably no aerosol content. However, this assumption is not strictly true. Previous studies [199] have shown that a residual aerosol contamination on the reference nights cannot be excluded. A correction of the VAOD of the reference nights of $\Delta\text{VAOD} = 0.016$ is suggested. Repeating the reconstruction of showers using this correction yields a deeper X_{max} which increases with energy:

$$\delta\left(X_{\text{max}}\right)_{\text{VAOD, norm.}} = -58.6 + 3.32 \log E \text{ g/cm}^2. \tag{5.92}$$

The accuracy of this VAOD re-normalisation correction is not clear. Therefore, the reconstructed X_{max} are not corrected, but instead a one-sided systematic uncertainty is assigned to account for the VAOD re-normalisation. Equation (5.92) is only valid above 10^{18} eV and extrapolation below zero does not make physical sense. Therefore, the assigned systematic uncertainty in this analysis will be rewritten as

$$\delta\left(X_{\text{max}}\right)_{\text{VAOD, norm.}} = \begin{cases} -58.6 + 3.32 \log E \text{ g/cm}^2 & \log E > 17.65 \\ 0 \text{ g/cm}^2 & \text{otherwise.} \end{cases} \tag{5.93}$$

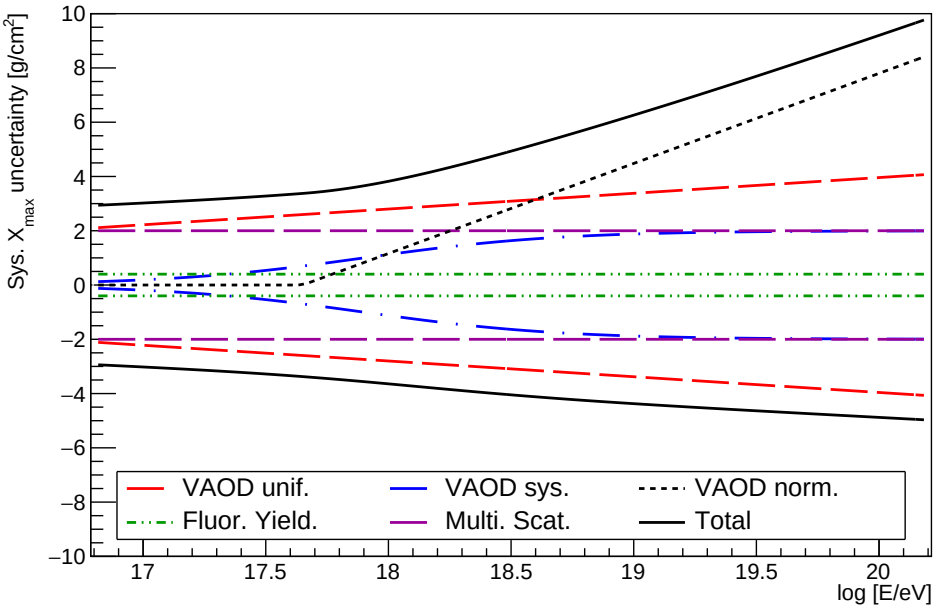


Figure 5.32. Summary of the systematic uncertainties of X_{max} due to atmospheric measurements.

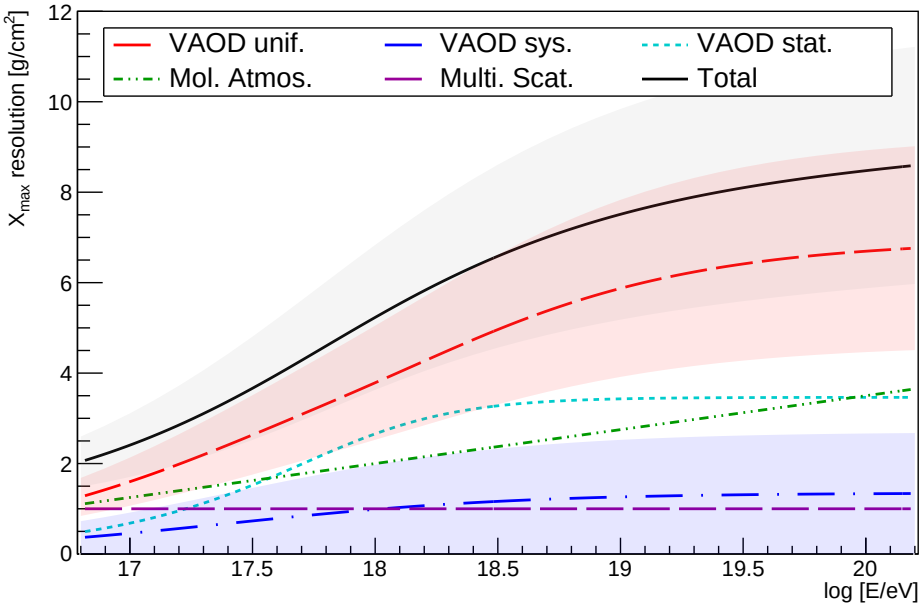


Figure 5.33. Individual components of the atmospheric contributions to the total $X_{\max}$ resolution. These components are all subtracted in quadrature from the calculated value of $\sigma\left(X_{\max}\right)$.

5.8.6 Summary

The following corrections are applied to the reconstructed X_{max} values

$$X_{\text{max}} = X_{\text{max}}^{\text{rec}} - \langle \Delta X_{\text{max}} \rangle - \frac{1}{2} \Delta \left(X_{\text{max}} \right)_{\text{align}} - b_{\text{LWcorr}} \tag{5.94}$$

where $X_{\text{max}}^{\text{rec}}$ is the reconstructed (biased) X_{max} value, $\langle \Delta X_{\text{max}} \rangle$ is the reconstruction bias, $\Delta \left(X_{\text{max}} \right)_{\text{align}}$ is the bias introduced by telescope mis-alignment, and b_{LWcorr} is the lateral width correction. The total systematic uncertainty of $\langle X_{\text{max}} \rangle$ is given by

$$\delta^2 \left(X_{\text{max}} \right) = \delta^2 \left(X_{\text{max}} \right)_{\text{rec}} + \delta^2 \left(X_{\text{max}} \right)_{\text{calib}} + \delta^2 \left(X_{\text{max}} \right)_{\text{accept}} + \delta^2 \left(X_{\text{max}} \right)_{\text{atmos}} \tag{5.95}$$

where

$$\begin{aligned} \delta^2 \left(X_{\text{max}} \right)_{\text{rec}} = {} & \delta^2 \left(X_{\text{max}} \right)_{\text{rec bias}} + \max\{ \delta^2 \left(X_{\text{max}} \right)_{\text{GH}} , \delta^2 \left(X_{\text{max}} \right)_{\text{constr}} \} \\ & + \delta^2 \left(X_{\text{max}} \right)_{\text{LWcorr}} \end{aligned} \tag{5.96}$$

$$\delta^2 \left(X_{\text{max}} \right)_{\text{calib}} = \delta^2 \left(X_{\text{max}} \right)_{\text{align}} + \delta^2 \left(X_{\text{max}} \right)_{\text{cal}} + \delta^2 \left(X_{\text{max}} \right)_{\delta t} \tag{5.97}$$

$$\begin{aligned} \delta^2 \left(X_{\text{max}} \right)_{\text{accept}} = {} & \delta^2 \left(X_{\text{max}} \right)_{\text{energy scale}} + \delta^2 \left(X_{\text{max}} \right)_{\text{fit uncert.}} \\ & + \delta^2 \left(X_{\text{max}} \right)_{\text{smooth}} + \delta^2 \left(X_{\text{max}} \right)_{\text{acc. corr.}} \end{aligned} \tag{5.98}$$

$$\begin{aligned} \delta^2 \left(X_{\text{max}} \right)_{\text{atmos}} = {} & \delta^2 \left(X_{\text{max}} \right)_{\text{FY}} + \delta^2 \left(X_{\text{max}} \right)_{\text{ms}} + \delta^2 \left(X_{\text{max}} \right)_{\text{VAOD, unif.}} \\ & + \delta^2 \left(X_{\text{max}} \right)_{\text{VAOD, sys.}} + \delta^2 \left(X_{\text{max}} \right)_{\text{VAOD, norm.}} \end{aligned} \tag{5.99}$$

These systematic uncertainties are summarised in Figure 5.34.

To compare the calculated $\sigma \left(X_{\text{max}} \right)$ to the predictions of hadronic models, the measured spread of the X_{max} distributions needs to be corrected for the total X_{max} resolution, including detector, calibration and atmospheric effects. The total X_{max} resolution, $\sigma \left(X_{\text{max}} \right)_{\text{total res.}}$ is given by

$$\sigma^2 \left(X_{\text{max}} \right)_{\text{total res.}} = \sigma^2 \left(X_{\text{max}} \right)_{\text{det}} + \sigma^2 \left(X_{\text{max}} \right)_{\text{atmos. res.}} \tag{5.100}$$

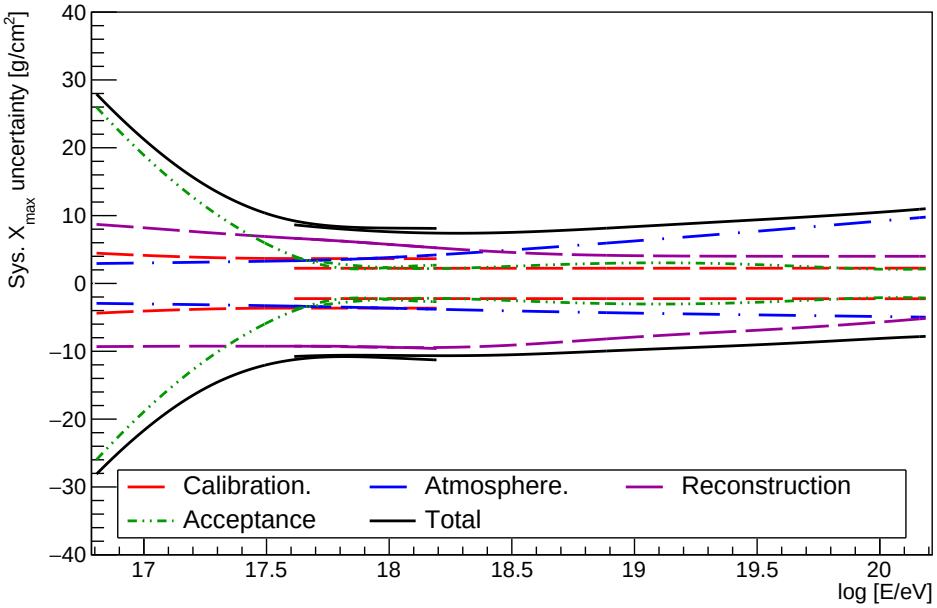


Figure 5.34. Summary of the $\langle X_{\text{max}} \rangle$ systematic uncertainties. Note the good agreement between the HeCo data set and the standard data set in the overlapping energy region ($17.6 \leq \log E \leq 18.2$).

where

$$\sigma^2 \left(X_{\text{max}}\right)_{\text{det}} = \sigma^2 \left(X_{\text{max}}\right)_{\text{det res.}} + \frac{1}{2}\sigma^2 \left(X_{\text{max}}\right)_{\text{align}} + \sigma^2 \left(X_{\text{max}}\right)_{\text{HeCo calib.}} \tag{5.101}$$

$$\begin{aligned}\sigma^2 \left(X_{\text{max}}\right)_{\text{atmos. res.}} = {} & \sigma^2 \left(X_{\text{max}}\right)_{\text{molAtmos}} + \sigma^2 \left(X_{\text{max}}\right)_{\text{ms}} + \sigma^2 \left(X_{\text{max}}\right)_{\text{VAOD, stat.}} \\ & + \frac{1}{2}\sigma^2 \left(X_{\text{max}}\right)_{\text{VAOD, sys.}} + \frac{3}{4}\sigma^2 \left(X_{\text{max}}\right)_{\text{VAOD, unif.}}\end{aligned} \tag{5.102}$$

As described in Section 5.4, the total resolution is subtracted in quadrature from the measured $\sigma \left(X_{\text{max}}\right)$. The components of the total detector resolution are summarised in Figure 5.35.

Finally, the total systematic uncertainty in $\sigma \left(X_{\text{max}}\right)$ is given by

$$\begin{aligned}\delta^2 \left[\sigma \left(X_{\text{max}}\right)\right] = {} & \delta^2 \left[\sigma \left(X_{\text{max}}\right)\right]_{\text{det.}} + \delta^2 \left[\sigma \left(X_{\text{max}}\right)\right]_{\text{end-to-end}} \\ & + \delta^2 \left[\sigma \left(X_{\text{max}}\right)\right]_{\text{accept}} + \delta^2 \left[\sigma \left(X_{\text{max}}\right)\right]_{\text{atmos.}}\end{aligned} \tag{5.103}$$

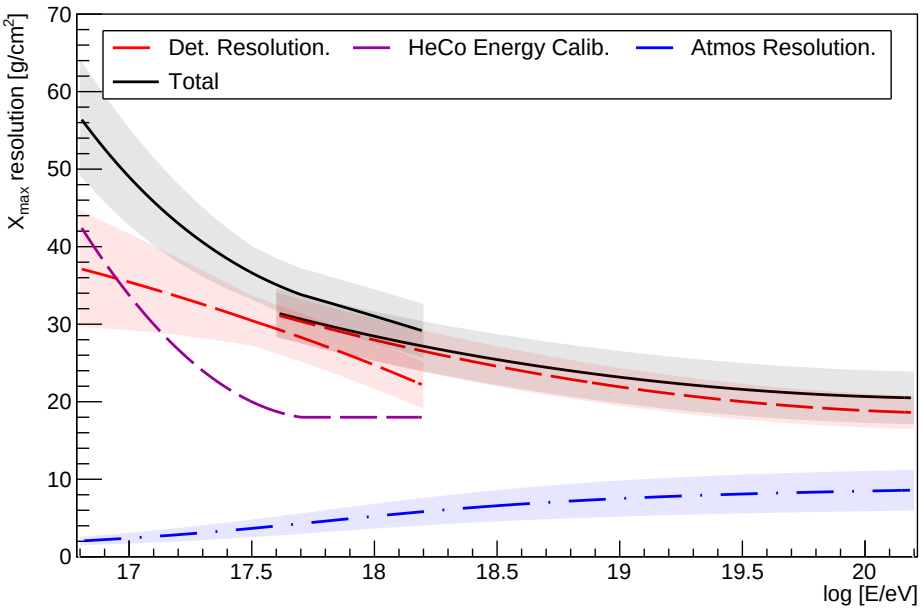


Figure 5.35. Summary of the total $X_{\max}$ resolution for both the HeCo and standard FD data sets.

where

$$\delta^2 \left[\sigma \left(X_{\max}\right)\right]_{\text{det.}} = \frac{1}{2}\delta^2 \left(X_{\max}\right)_{\text{align}} \tag{5.104}$$

$$\begin{aligned}\delta^2 \left[\sigma \left(X_{\max}\right)\right]_{\text{accept}} = \delta^2 \left[\sigma \left(X_{\max}\right)\right]_{\text{energy scale}} + \delta^2 \left[\sigma \left(X_{\max}\right)\right]_{\text{fit uncert.}} \\ + \delta^2 \left[\sigma \left(X_{\max}\right)\right]_{\text{smooth}} + \delta^2 \left[\sigma \left(X_{\max}\right)\right]_{\text{acc. corr.}}\end{aligned} \tag{5.105}$$

$$\delta^2 \left[\sigma \left(X_{\max}\right)\right]_{\text{atmos.}} = \frac{1}{2}\sigma^2 \left(X_{\max}\right)_{\text{VAOD, sys.}} + \frac{1}{4}\sigma^2 \left(X_{\max}\right)_{\text{VAOD, unif.}} \tag{5.106}$$

(5.107)

The systematic uncertainties in $\sigma \left(X_{\max}\right)$ are summarised in Figure 5.36.

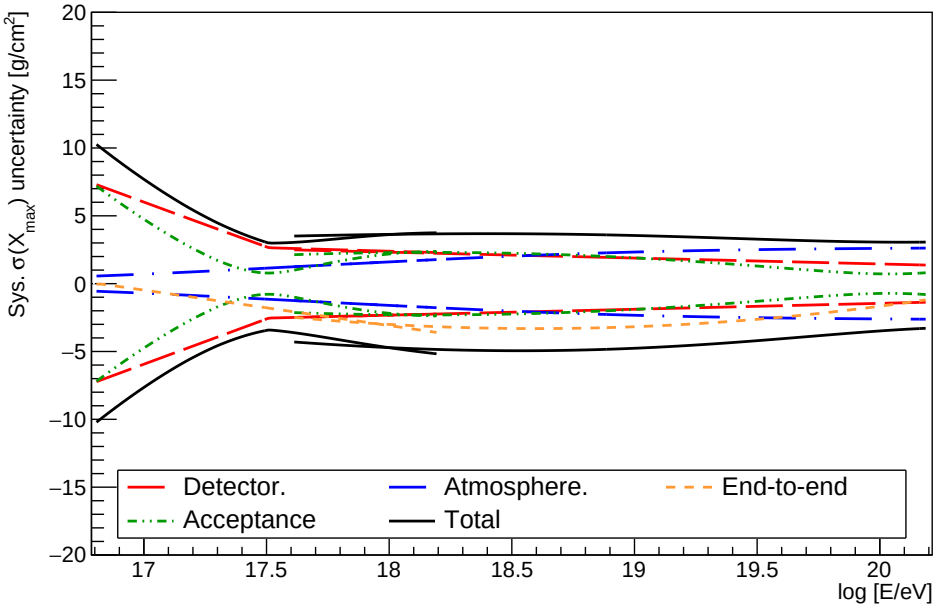


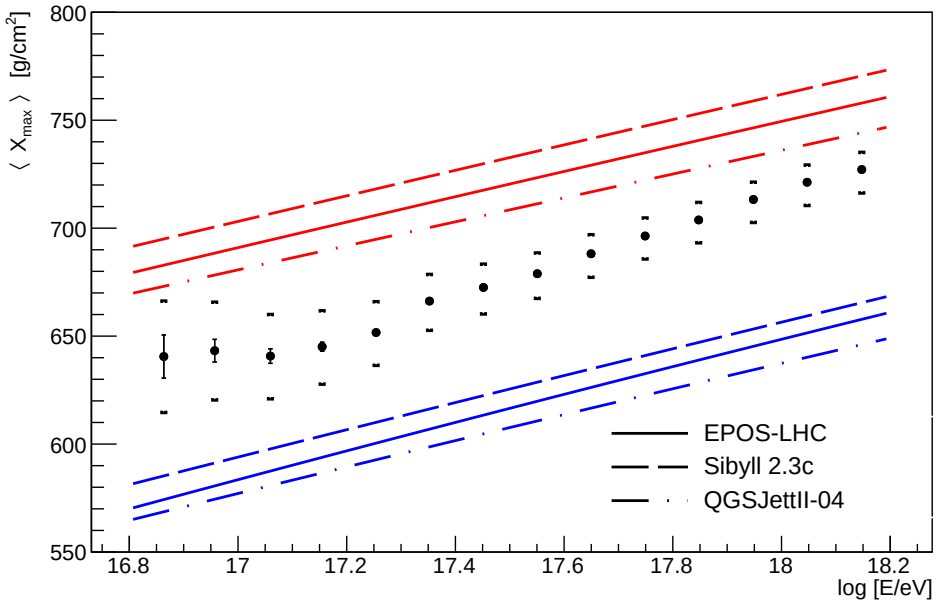
Figure 5.36. Summary of the $\sigma\left(X_{\max}\right)$ systematic uncertainties. Note the good agreement between the HeCo data set and the standard data set in the overlapping energy region ($17.6 \leq \log E \leq 18.2$).

5.9 X_{max} Data Analysis

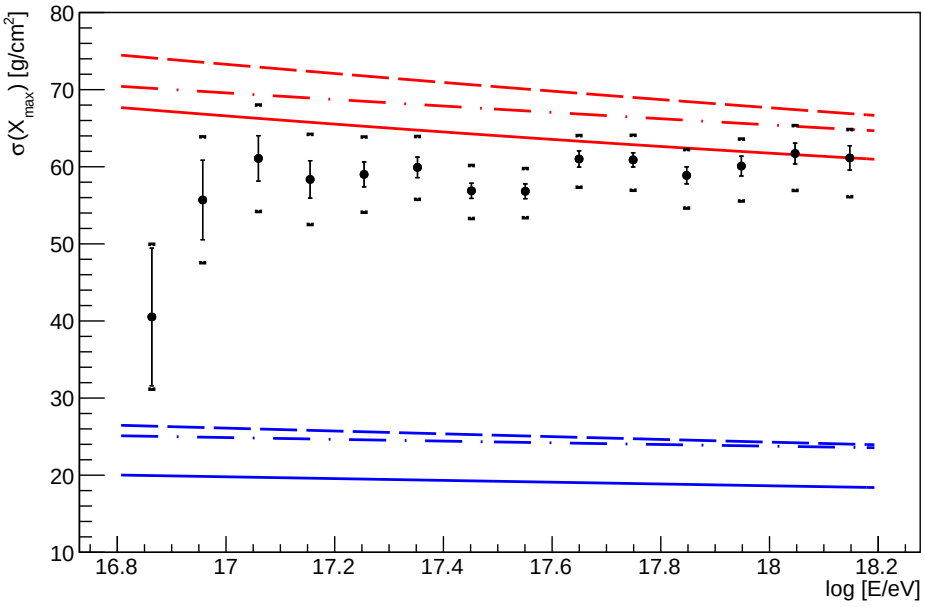
Following the proof-of-concept of the kernel acceptance method with simulated data sets in Section 5.7, the newly developed analysis technique is applied to the measured HeCo and standard FD data sets. All selection cuts have been applied as described in Section 5.5. The X_{max} acceptance is estimated using the kernel estimation model described in Section 5.3 in lieu of the FoV cuts and subsequent Λ_η acceptance model. The kernel acceptance model is applied in all energy bins from $10^{16.8}$ to $10^{20.2}$ eV in order to compare the calculated X_{max} moments to the predictions of hadronic models. The calculated moments are compared between the HeCo analysis procedure between $10^{17.2}$ eV and $10^{18.1}$ eV, and the standard FD analysis above $10^{17.6}$ eV as well as to the moments calculated using the standard analysis technique in [90].

5.9.1 X_{max} Moments

The final X_{max} moments calculated using the kernel acceptance, including all corrections and systematic uncertainties, are presented in this section. Shown in Figure 5.37 are the X_{max} moments calculated for the HeCo data set compared with the predictions from three hadronic models; EPOS-LHC, Sibyll 2.3c, and QGSJetII-04. The results include four additional energy bins at the lowest energies between $10^{16.8}$ eV and $10^{17.2}$ eV compared to the previous analysis. There is a hint of a break in the elongation rate at $\sim 10^{17}$ eV which may be associated with the change in the energy spectrum at the second knee. Such a result would potentially provide insight to the mechanism from which the second knee arises. However, the indication from $\sigma\left(X_{\text{max}}\right)$ is that the composition is still getting heavier below $10^{17.2}$ eV, contradicting the lighter mass indicated by $\langle X_{\text{max}} \rangle$. The two lowest energy data points in Figure 5.37a have large statistical uncertainties. Therefore, a change in composition is not statistically significant. The moments for the standard FD data set is also shown in Figure 5.38. The results include an additional energy bin at the highest energies $\left(\geq 10^{19.7} \text{ eV}\right)$. Both moments suggest a heavy component to the mass composition at the highest energies, which is consistent with previous results from Auger [90]. There is also some evidence for a flattening of $\sigma\left(X_{\text{max}}\right)$ at the highest energies, where the composition may have reached its heaviest. This may give some insight to which of the hadronic interaction models best represent nature.

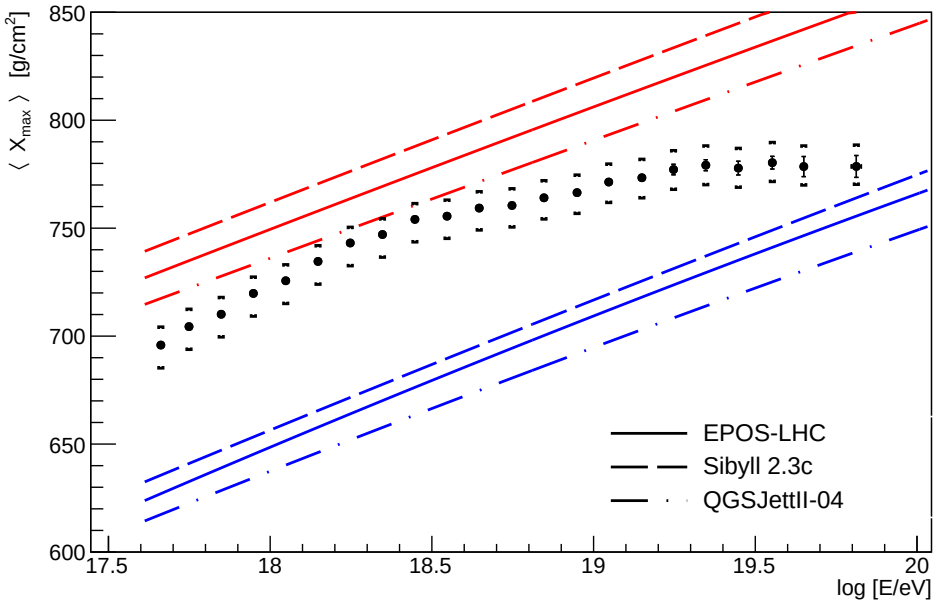


(a)

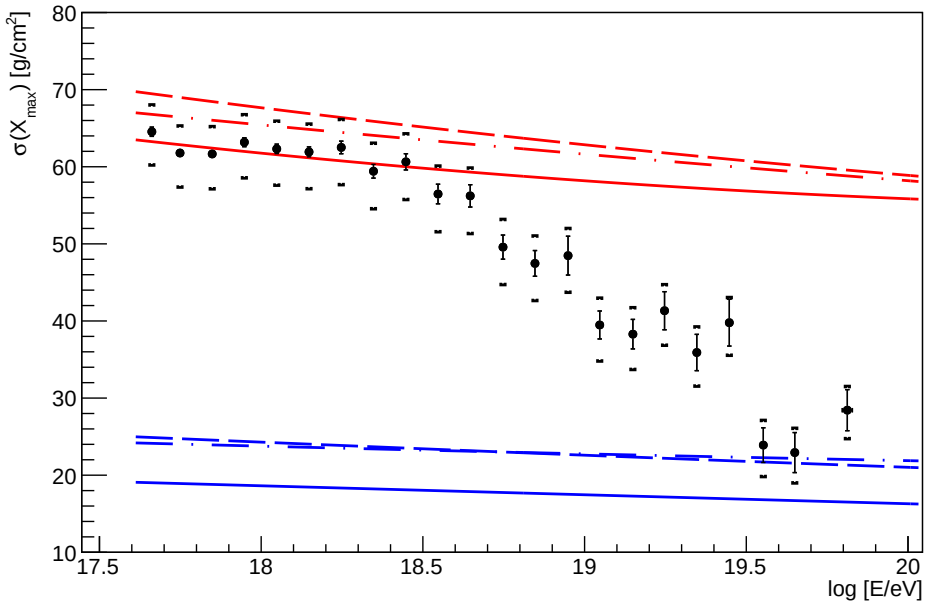


(b)

Figure 5.37. The moments of the $X_{\max}$ distributions from the HeCo data set. The moments are compared to the prediction of the hadronic models, EPOS-LHC (solid lines), QGSJetII-04 (dash-dot lines) and Sibyll 2.3c (dashed lines). Systematic uncertainties are indicated by the brackets for each point.



(a)



(b)

Figure 5.38. The moments of the $X_{\max}$ distributions from the standard FD data set. The moments are compared to the prediction of the hadronic models, EPOS-LHC (solid lines), QGSJetII-04 (dash-dot lines) and Sibyll 2.3c (dashed lines). Systematic uncertainties are indicated by the brackets for each point.

5.9.2 Combining the HeCo and Standard FD Moments

The calculated X_{max} moments from the HeCo and standard FD data sets are compared in Figure 5.39. The overlapping energy range of $17.6 < \log E < 18.2$ provides a region where the moments can be compared directly. The comparison appears to show small systematic differences between the two data sets; the X_{max} moments of the HeCo data set are systematically lower than those of the standard FD moments. This is not very surprising considering that there are known calibration differences which so far, remain unaccounted for. This difference has also been observed in the standard X_{max} analysis with fiducial FoV cuts. This is still under investigation by Auger. The same procedure used in the standard analysis will be adopted to combine the HeCo and standard FD moments. That is, the HeCo moments are shifted to agree with the standard FD moments in the overlapping energy range. In order to derive the shift required to make the HeCo and standard FD moments agree in the overlapping region, the difference between the moments are plotted as a function of energy in Figure 5.40. Note that the differences are constant with energy for both $\langle X_{\text{max}} \rangle$ and $\sigma\left(X_{\text{max}}\right)$. The fitted shift for $\langle X_{\text{max}} \rangle$ is

$$\Delta \langle X_{\text{max}} \rangle = (6.87 \pm 0.67) \text{ g/cm}^2, \tag{5.108}$$

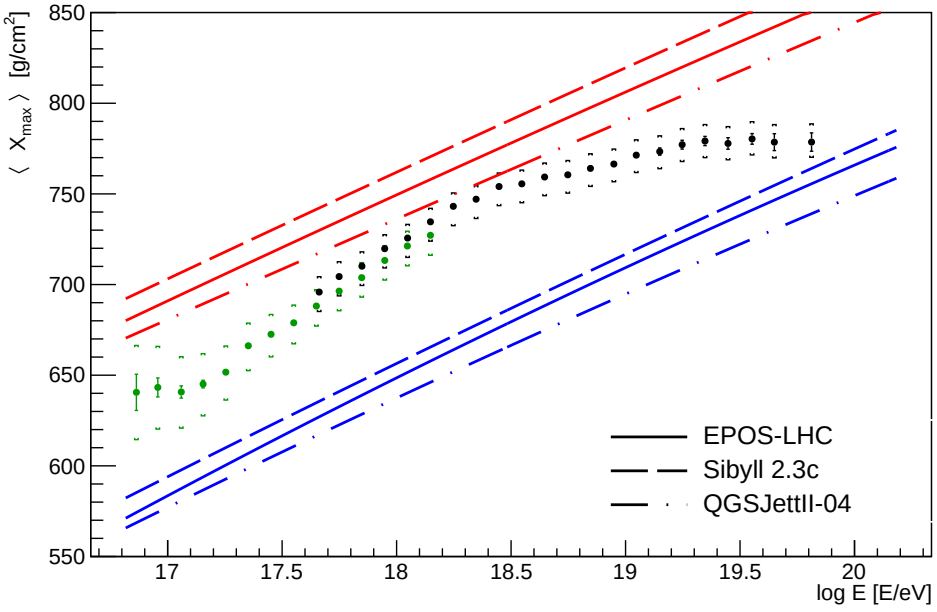
and the shift for $\sigma\left(X_{\text{max}}\right)$ is

$$\Delta\sigma\left(X_{\text{max}}\right) = (2.0 \pm 0.71) \text{ g/cm}^2. \tag{5.109}$$

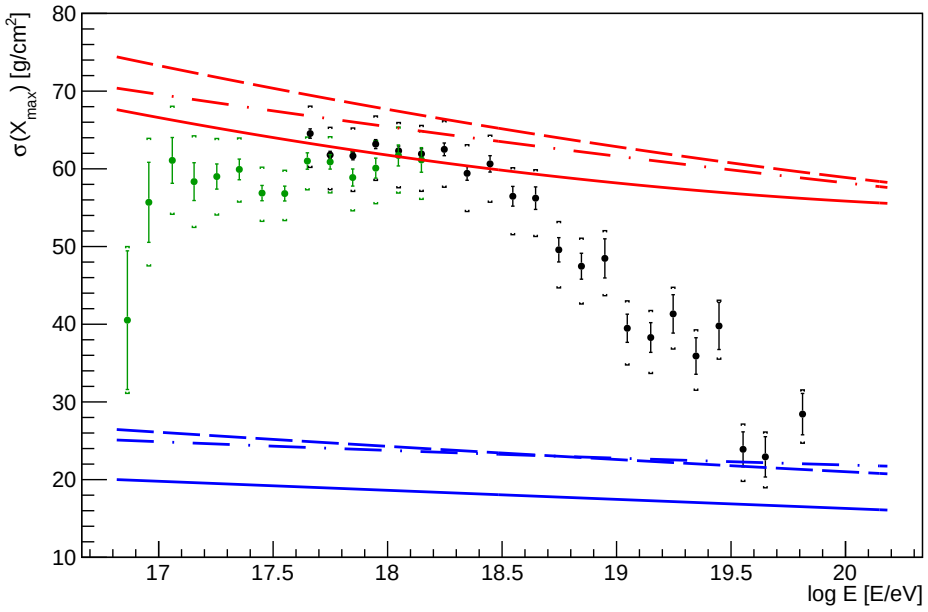
The shifted HeCo moments are compared with the standard FD moments in Figure 5.41. With the shift applied to the HeCo moments, the two data sets now show close agreement in both magnitude and shape. The HeCo and standard FD moments are combined in the overlapping energy range by calculating the weighted mean and variance of the HeCo and standard FD moments using the equations described in Section 5.4. However, the weights in this case are instead given by the inverse squared error of the moments. That is

$$w_i = \frac{1}{\sigma_i^2} \tag{5.110}$$

where σ_i is the uncertainty in $\langle X_{\text{max}} \rangle$ and $\sigma\left(X_{\text{max}}\right)$ for the individual HeCo and standard FD data sets. Since the HeCo moments are shifted to match the standard FD moments, the systematic uncertainties in the overlapping energy bins are assigned according to the standard FD systematic uncertainties. However, the systematic uncertainties of the two datasets are almost identical in the overlapping energy bins, as shown in Figure 5.34. The final combined moments are shown in Figure 5.42.

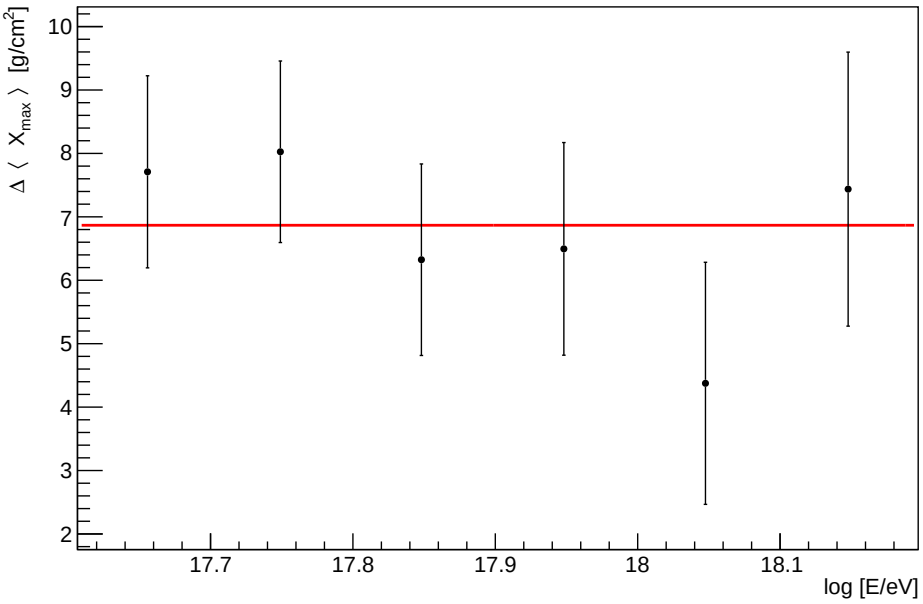


(a)

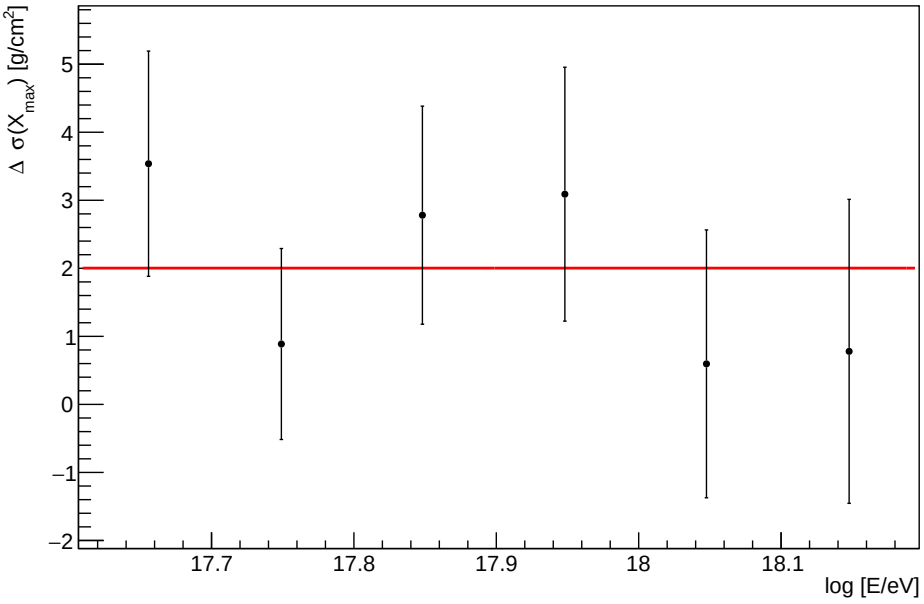


(b)

Figure 5.39. A comparison of the moments of the $X_{\max}$ distributions between the HeCo (green) and the standard FD (black) data sets. The moments are compared to the prediction of the hadronic models, EPOS-LHC (solid lines), QGSJetII-04 (dash-dot lines) and Sibyll 2.3c (dashed lines).

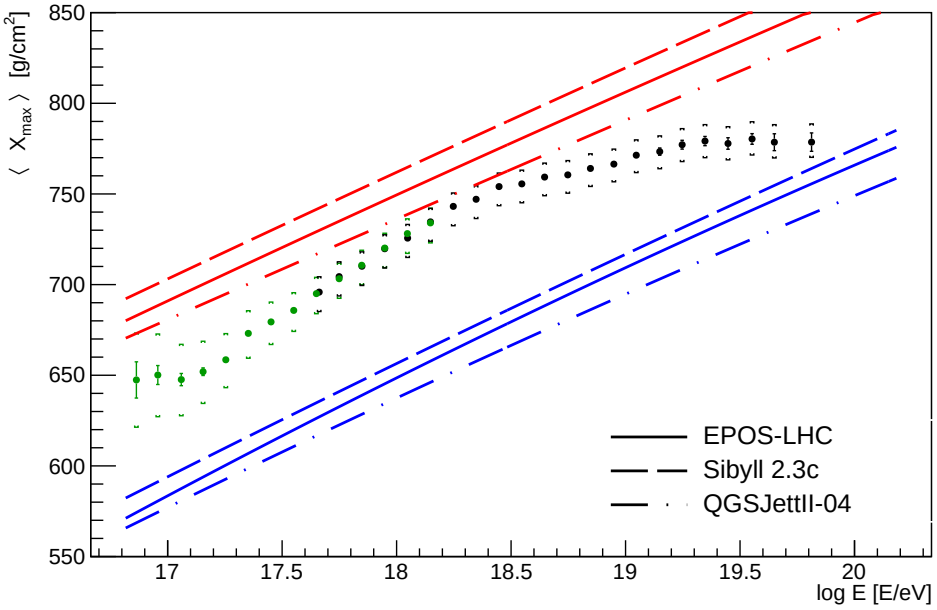


(a)

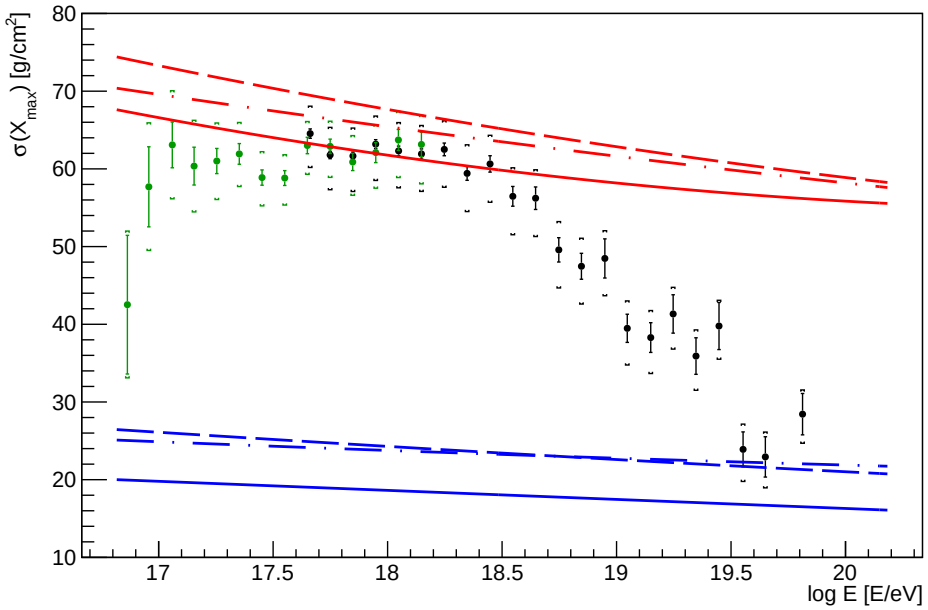


(b)

Figure 5.40. The difference between the moments of the $X_{\max}$ distributions between the HeCo and the standard FD data sets in the overlapping energy bins. The difference is constant in energy, represented by the fitted red line.

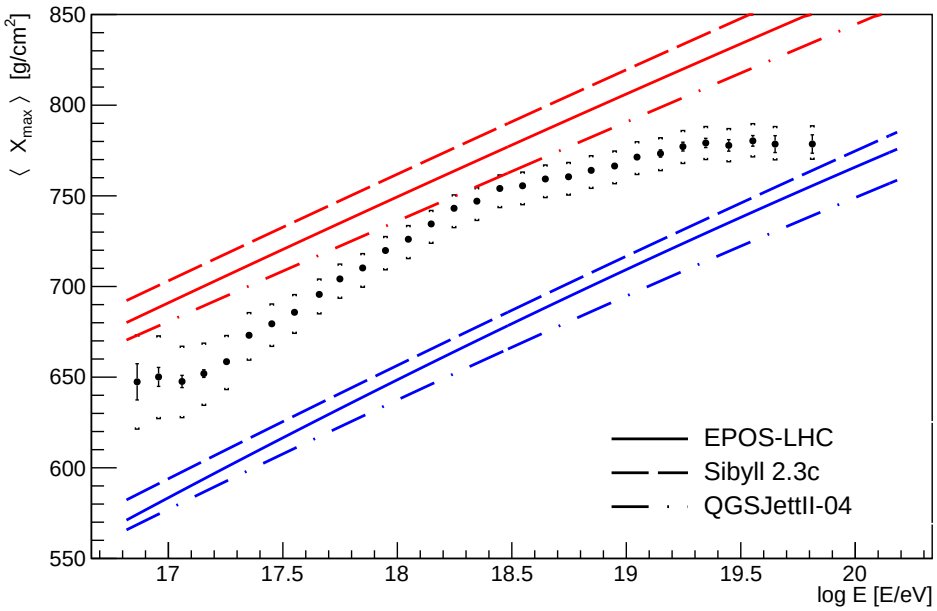


(a)

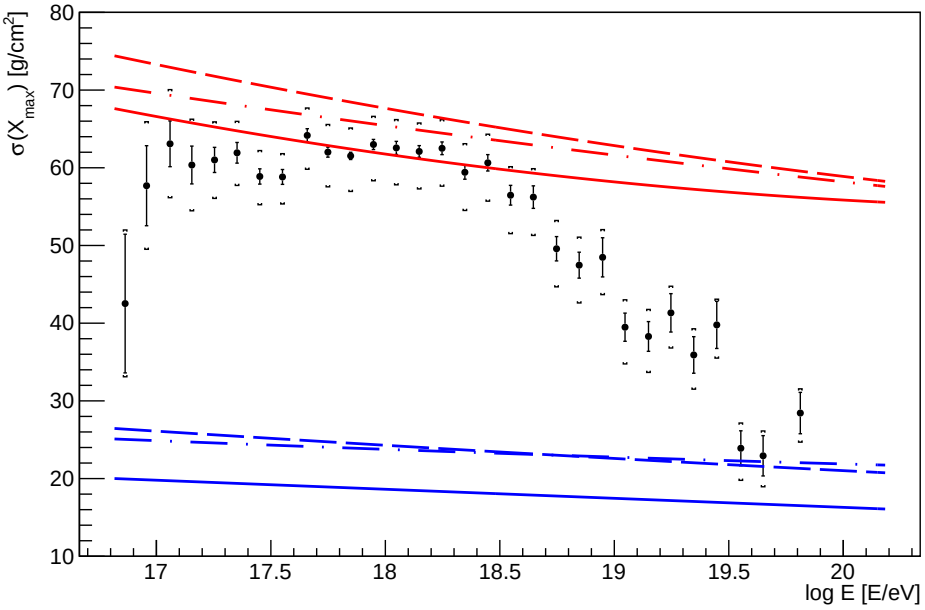


(b)

Figure 5.41. A comparison of the moments of the $X_{\max}$ distributions between the HeCo (green) and the standard FD (black) data sets with the HeCo moments shifted by the average difference in the overlapping energy bins. The moments are compared to the prediction of the hadronic models, EPOS-LHC (solid lines), QGSJetII-04 (dash-dot lines) and Sibyll 2.3c (dashed lines).



(a)



(b)

Figure 5.42. The combined moments of the $X_{\max}$ distributions. The moments are compared to the prediction of the hadronic models, EPOS-LHC (solid lines), QGSJetII-04 (dash-dot lines) and Sibyll 2.3c (dashed lines).

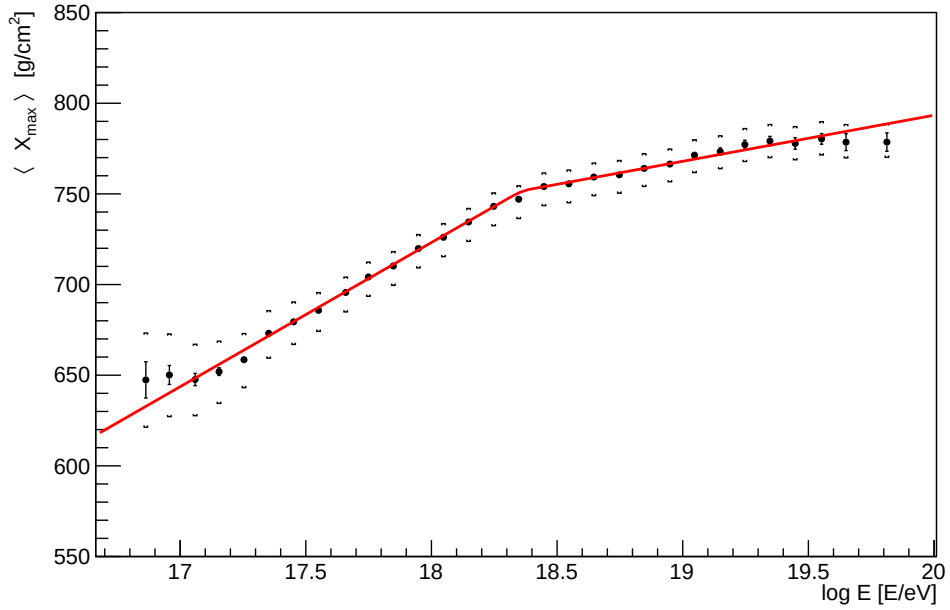


Figure 5.43. The combined $\langle X_{\text{max}} \rangle$ measurements and the fit (red line) of the elongation rate, assuming a single-broken line.

5.9.3 Elongation Rate

As discussed in Sections 2.4 and 2.5, the elongation rate is defined as the rate of change in $\langle X_{\text{max}} \rangle$ per decade of energy

$$\Lambda = \frac{d \langle X_{\text{max}} \rangle}{d \log_{10} E}. \tag{5.111}$$

The elongation rate is one of the most reliable measurements of mass composition, since it is almost independent of hadronic models. It is clear from Figure 5.42a that the data is compatible with pure proton at $\sim 10^{18.3}$ eV and becomes heavier in composition at both lower and higher energies with an approximately linear change in X_{max}. Based on this observation a single-broken line is fit to the data in order to determine the elongation rate above and below the break at $\sim 10^{18.3}$ eV. The fit is shown in Figure 5.43. The fit function has the form

$$\langle X_{\text{max}} \rangle (\log E) = \langle X_{\text{max}} \rangle_0 + \begin{cases} \Lambda_1 (\log E - \log E_0) & \log E < \log E_0 \\ \Lambda_2 (\log E - \log E_0) & \log E \geq \log E_0 \end{cases} \tag{5.112}$$

The best fit parameters are given by

$$\Lambda_1 = 79.6 \pm 0.7|_{\text{stat}}\, ^{+8.1}_{-6.2}|_{\text{syst}}\ \text{g/cm}^2 \tag{5.113}$$

$$\Lambda_2 = 25.4 \pm 1.5|_{\text{stat}}\, ^{+5.2}_{-0.6}|_{\text{syst}}\ \text{g/cm}^2 \tag{5.114}$$

$$\log E_0/\text{eV} = 18.36 \pm 0.02|_{\text{stat}}\, ^{+0.05}_{-0.1}|_{\text{syst}} \tag{5.115}$$

The systematic uncertainties are determined by re-fitting the elongation rates after shifting $\langle X_{\text{max}} \rangle$ up and down by their systematic uncertainty. This involves shifting $\langle X_{\text{max}} \rangle$ up and down independently by each component of the total systematic uncertainty. The four components of the total systematic uncertainty are shown in Figure 5.34. The components are calibration, atmosphere, reconstruction, and acceptance. Since the elongation rate is defined by the slope of $\langle X_{\text{max}} \rangle$ as function of energy, only the systematic uncertainties which are correlated with energy are expected to significantly affect the elongation rate. Nevertheless, the systematic uncertainties in the three parameters of the elongation rate fit are calculated for each component of the total $\langle X_{\text{max}} \rangle$ systematic uncertainty, and the results are summarised in Table 5.6. The total is determined by adding the systematic uncertainties of each component in quadrature. The largest source of systematic uncertainty is the acceptance correction, especially for the elongation rate below the break. This is expected since the acceptance correction contributed the most to the total $\langle X_{\text{max}} \rangle$ systematic uncertainty in the HeCo energy range. The atmosphere contribution is smaller at lower energies, and larger at higher energies where the energy dependence of the atmosphere component of the total $\langle X_{\text{max}} \rangle$ systematic uncertainty is strongest. The contributions from the reconstruction and calibration systematics are small, but not negligible due to their weak correlation with energy.

	Λ_1 [g/cm^2]	Λ_2 [g/cm^2]	$\log E_0$/eV
Calibration	$+2.3$ -0.2	$+2.9$ -0	$+0$ -0.06
Atmosphere	$+1.4$ -0.8	$+2.7$ -0.6	$+0.004$ -0.001
Reconstruction	$+0$ -1.4	$+2.4$ -0	$+0.003$ -0.05
Acceptance	$+7.7$ -6.0	$+2.3$ -0	$+0.05$ -0.09
Total	$+8.1$ -6.2	$+5.2$ -0.6	$+0.05$ -0.1

Table 5.6. The systematic uncertainties of the elongation rate fit parameters for each component of the total $\langle X_{\text{max}} \rangle$ systematic uncertainty. The total systematic uncertainty of each parameter is determined by adding the individual components in quadrature.

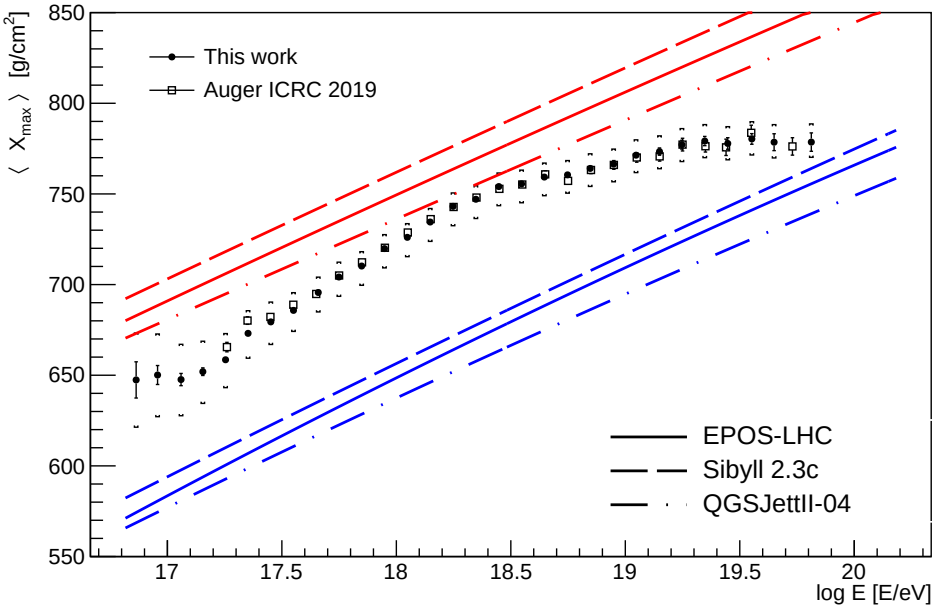
Very similar results were determined in the ICRC 2019 mass composition analysis. The break in the elongation rate is observed at $\log E_0/\text{eV} = 18.32 \pm 0.03$ and the elongation rate is 77 ± 2 g/cm^2/dec, and 26 ± 2 g/cm^2/dec below and above the break, respectively [90]. Therefore, the results of the two analyses are compatible within statistical uncertainty.

5.9.4 Comparison with Standard Analysis

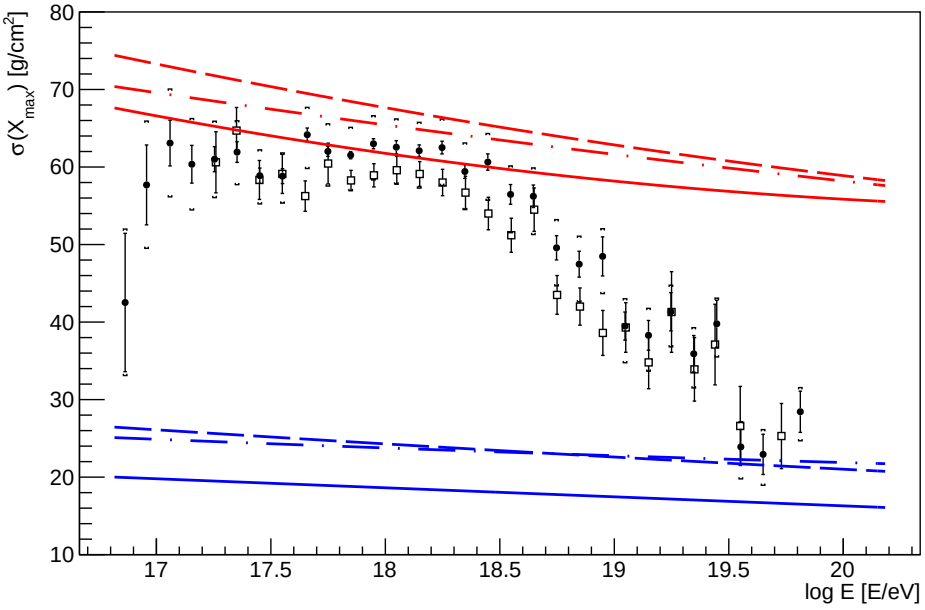
In order to further validate the kernel acceptance model and the relaxed data selection cuts, the combined X_{max} moments derived in this work, are compared with the ICRC 2019 X_{max} moments [90] in Figure 5.44. There is very good agreement between the moments derived in this work and the standard analysis. There is a small divergence in the moments in the lowest three energy bins of the ICRC results, however the shape of $\langle X_{\text{max}} \rangle$ as a function of energy is almost identical. The additional energy bin at the highest energies continues the pattern of the previous analysis, providing further evidence for a heavy component of the mass composition at the highest energies. The agreement with the previous analysis further validates the X_{max} moments derived using the kernel acceptance model and without FoV cuts. The number of events passing the selections cuts described in Section 5.5 are compared to that of the ICRC 2019 analysis in Figure 5.45. The modified selection cuts utilised in this work increases the number of events in most energy bins by at least a factor of 2. The total number of events in all energy bins in this work is 156,717 compared to 47,863 in the ICRC 2019 analysis. The additional energy bins included in this work are only possible due to the increased number of events which have been included in the analysis as a result of removing the fiducial FoV cuts. There is a small systematic difference of ~ 3 g/cm^2 between $\sigma\left(X_{\text{max}}\right)$ of the two analyses above $\sim 10^{18}$ eV. This can be explained by the bias correction applied to $\sigma\left(X_{\text{max}}\right)$ to account for the X_{max} dependent reconstruction bias as described in Section 5.7.2. As discussed in Section 5.7.2, the correction should also be applied to the standard analysis. This would bring the results into agreement.

5.9.5 Comparison with Other Experiments

A comparison between the X_{max} moments determined in this work with those determined by other experiments is shown in Figure 5.46. The results are compared with Yakutsk [211], Tunka [212], the Low Frequency Array (LOFAR) [213] and the Telescope Array (TA) [214]. Also shown for reference are the Auger results from ICRC 2019 [90]. Generally good agreement with other experiments is observed. There are, however, some notable differences. At lower energies the Auger $\langle X_{\text{max}} \rangle$ results (both from this work and the standard analysis) seem to diverge from the



(a)



(b)

Figure 5.44. The combined moments of the $X_{\max}$ distributions derived in this work using the kernel acceptance method, compared to the ICRC 2019 $X_{\max}$ moments [90] which were derived using the standard $X_{\max}$ analysis. The agreement is extremely good between the two analyses. The moments are compared to the prediction of the hadronic models, EPOS-LHC (solid lines), QGSJetII-04 (dash-dot lines) and Sibyll 2.3c (dashed lines).

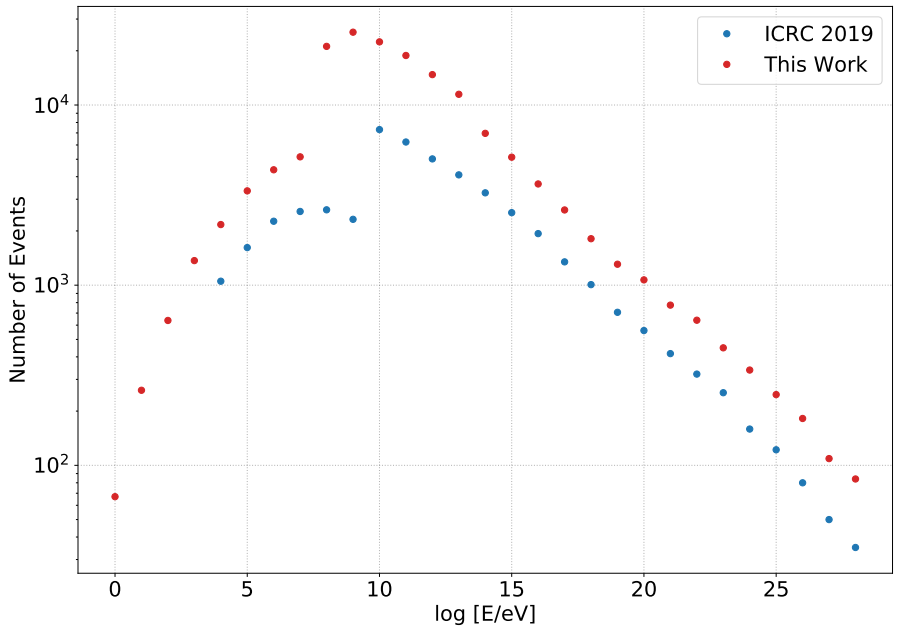


Figure 5.45. The number events passing selection cuts (described in Section 5.5) shown in red, compared with the number of events passing selection cuts in the ICRC 2019 data set [90] shown in blue. Note the significant increase in the number of events passing selection cuts in this work compared to the analysis for ICRC 2019.

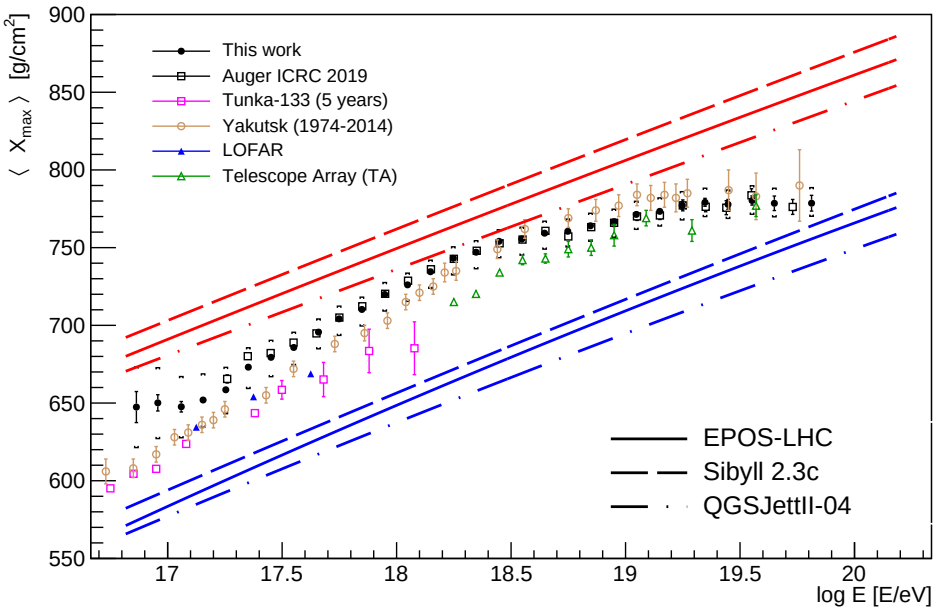
other experiments, although the results remain within systematic uncertainties. An important distinction between Auger and Yakutsk, Tunka, and LOFAR is that the Auger results are based on fluorescence measurements, while the others are based on Cherenkov light or radio measurements. It is not clear why such a difference would exist and this warrants further investigation. The best agreement between Auger measurements and Yakutsk is around the energy break at $\sim 10^{18.4}$ eV. At the highest energies, Auger dominates in regard to statistical reliability with small uncertainties. The large statistical uncertainties of the Yakutsk data mean that the results overlap with the Auger results. However, there does seem to be a distinct systematic divergence above the energy break.

5.10 Conclusions

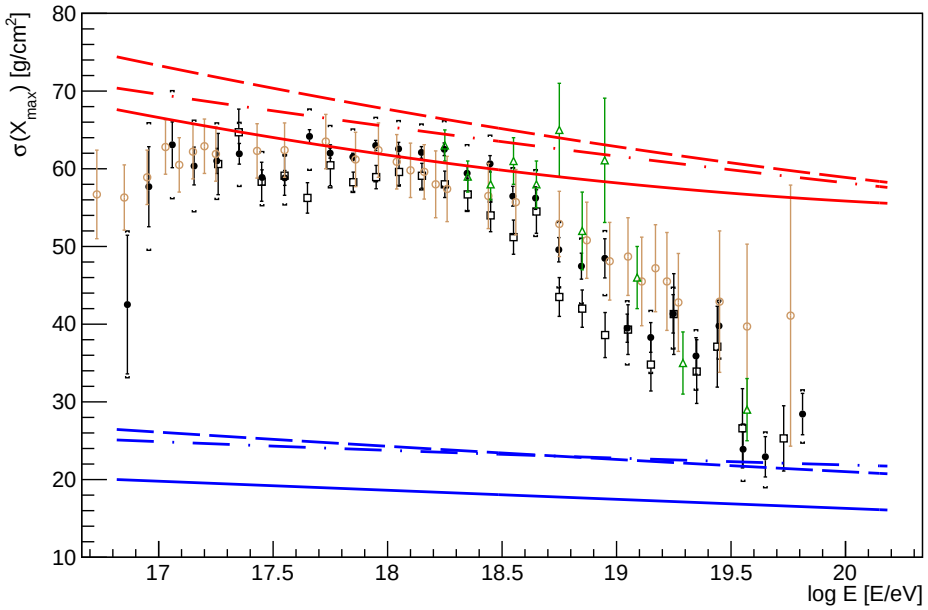
It has been shown that using a kernel estimation model in conjunction with flat X_{max} RealMC simulations, the X_{max} detector acceptance can be derived without applying the standard FoV cuts. Very good agreement in the calculated X_{max} moments is observed between the results from this analysis and the results from ICRC 2019 which used the standard analysis including the FoV selection cuts and the Λ_η acceptance correction. The close agreement between the two analyses validates the kernel acceptance model.

While additional data can be included in the X_{max} analysis with relaxed cuts, one downside is the increased systematic uncertainties due to the acceptance correction compared to the standard analysis. The balance between statistical and systematic uncertainties should be considered in the choice between the two analyses. The biggest benefit of the kernel acceptance analysis developed in this work is the extension of the energy range of the X_{max} analysis in both directions; the energy threshold of the HeCo analysis was reduced from $10^{17.2}$ eV to $10^{16.8}$ eV, and the standard FD analysis adds an additional energy bin at the highest energies. With future additional data at these extreme energies, significant astrophysical implications may be uncovered by mass composition analysis of UHECRs.

Further investigation of the kernel acceptance method could include cross-checks on various subsets of the data such as vertical and inclined events and independent results for each FD. The apparent shift in the $\langle X_{\text{max}} \rangle$ between fluorescence and other measurement techniques warrants additional investigation. An understanding of the differences will be important for accurate measurements of cosmic ray mass composition over the full energy range and to understand the features of the energy spectrum.



(a)



(b)

Figure 5.46. The $X_{\max}$ moments calculated in this analysis for the HeCo data set (solid green) and the standard FD data set (solid black) compared with the results of other experiments.

6 The Fluorescence Detector Array of Single-Pixel Telescopes

The cosmic ray flux above the suppression is less than one per century per square kilometre. To account for this low flux, detectors must span a very large area in order to measure significant numbers of the highest energy cosmic rays. The Fluorescence detector Array of Single-pixel Telescopes (FAST) is a design concept for the next generation of ultra-high energy cosmic ray (UHECR) observatories [215–217]. It addresses the requirements for a large-area, inexpensive detector capable of measuring cosmic ray properties at the highest energies, while maintaining adequate energy, X_{max} and angular resolution [215, 217, 218]. FAST is a compact and portable fluorescence telescope. Central to its design is a camera consisting of only four 200 mm photomultiplier tubes (PMTs). In order to achieve the desired exposure to the highest energy cosmic rays, FAST telescopes can be arranged to form a large array.

6.1 Concept and Prototype Telescope Design

The FAST design, consists of a telescope with a field of view (FoV) of $30^\circ \times 30^\circ$, covered by a camera of only four PMTs located at the focal plane of a mirror with a 1 m^2 collecting area [218]. This modification with respect to the design of the Pierre Auger Observatory (Auger) Fluorescence Detector (FD) telescopes, represents a significant reduction in the detector cost. This allows many FAST telescopes to be deployed over a very large area. The FAST concept consists of an array of FD stations, each consisting of 12 co-located telescopes which, in total, provide a 360° azimuthal FoV. These stations would be arranged in a triangular grid much like the Surface Detector (SD) of Auger, but with a significantly larger spacing of ~ 20 km. The spacing of the array can be optimised based on the desired energy threshold.

The FAST design and configuration come at the cost of reduced low energy performance due to the increased night-sky background (NSB) observed by such large PMTs. The signal-to-noise ratio of a PMT is proportional to $\sqrt{A/\Delta\Omega}$ [219], where $A \sim$1 m^2 is the collecting area of the telescope and $\Delta\Omega \sim 15^\circ \times 15^\circ$ is the pixel solid angle. In comparison, the FD telescopes from Auger have an effective collecting area $A \sim 3$ m^2 and pixel solid angle $\Delta\Omega \sim 1.5^\circ \times 1.5^\circ$ [136, 144]. As an addi-

Figure 6.1. Left: The FAST telescope design with mirror segments visible during installation. Centre: Arrangement of the four PMTs in the camera box. Right: The first installed FAST telescope inside a protective shelter at the TA BRM FD site.

tional constraint, the coarse granularity of the FAST camera provides insufficient information for geometrical reconstruction using standard analysis techniques such as that employed by Auger. Instead, FAST can provide an independent shower energy and $X_{\max}$ measurement if paired with the geometry obtained from a surface array. Alternatively, FAST can be operated in stereo mode using multiple FAST telescopes deployed as an array. For showers which trigger multiple FAST stations, their measured signals can also be combined to constrain the shower geometry. This is investigated in Chapters 7 to 9.

The first full-size prototype telescope was installed at the Black Rock Mesa (BRM) site at the Telescope Array (TA) in Utah, USA in October 2016 following successful testing of the FAST concept performed in 2014 [215, 220, 221]. Figure 6.1 shows the design of the telescope and PMT arrangement. An additional two FAST telescopes were installed at the TA site in 2017 and 2018, and another was installed at the Los Leones FD site of Auger in early 2019.

6.1.1 Telescope Design

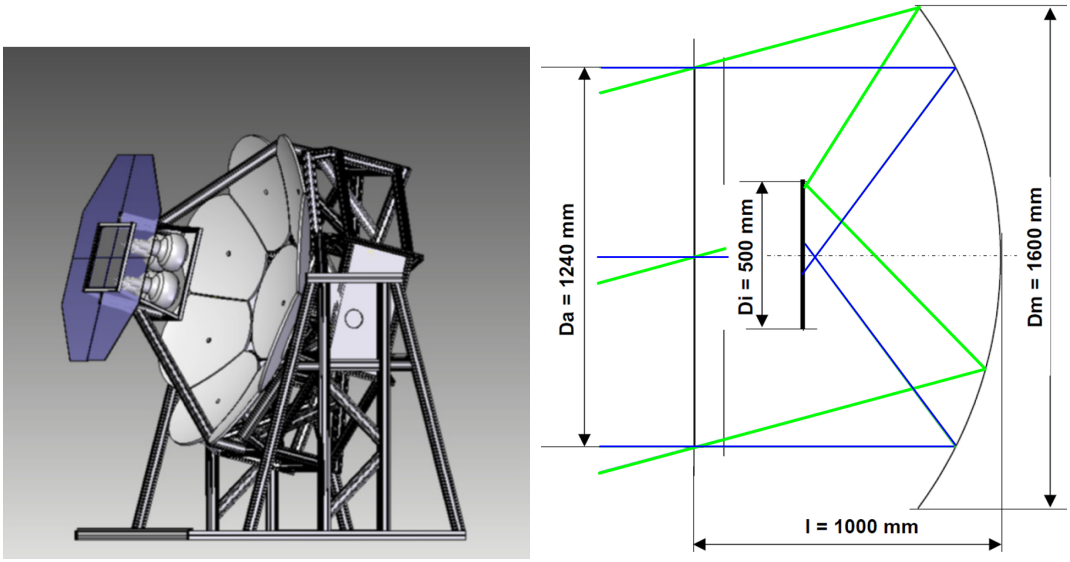
The full-size FAST prototype telescope utilises a Schmidt type optical design, with the camera comprising four 200 mm PMTs. Typically, a corrector plate is placed at the mirror's radius of curvature in a Schmidt design, at a distance of $2f$ (where f is the focal length of the mirror) in order to control off-axis aberrations, namely coma and astigmatism. However, for the FAST design, the telescope takes the form of a lens-less Schmidt camera since only four PMTs cover the large angular FoV of $\sim 30^\circ \times 30^\circ$. Therefore, the size and shape requirements of the point spread

function (PSF)[1] can be relaxed without affecting the functionality of the telescope. Consequently, the use of a corrector plate is unnecessary. The design of the telescope frame is shown in Figure 6.2a and the dimensions of the FAST optical system is shown in Figure 6.2b. The telescope mirror is reduced in size relative to the Auger FD telescopes to 1 m^2 and the distance from the mirror to the focal surface and entrance aperture is reduced, resulting in a more compact telescope compared to a regular Schmidt telescope. A schematic of the prototype is shown in Figure 6.2a and the FAST telescope dimensions are shown in Figure 6.2b. An octagonal aperture of width 1.24 m is positioned 1 m from a 1.6 m diameter segmented spherical mirror. The telescope has an effective collecting area of 1 m^2 (taking the camera and structure shadow into account) and a FoV of $30^\circ \times 30^\circ$ [218]. The segmented design of the mirror is the most practically convenient and least expensive design. The mirror consists of a central circular segment and 8 surrounding segments, or "petals" made from a borosilicate glass substrate. These are produced by the Joint Laboratory of Optics of the Palacký University and the Institute of Physics of the Academy of Sciences of the Czech Republic. The mirror reflectivity is almost constant in the fluorescence wavelength band of 300-420 nm as shown in Figure 6.3b. The aperture of the FAST telescopes incorporates an ultra-violet (UV) filter to reduce the NSB for photons with wavelengths above 400 nm. The filter transmittance as a function of wavelength is shown in Figure 6.3b. More detailed information regarding the FAST optical design can be found in [218, 222].

6.1.2 Optical Performance

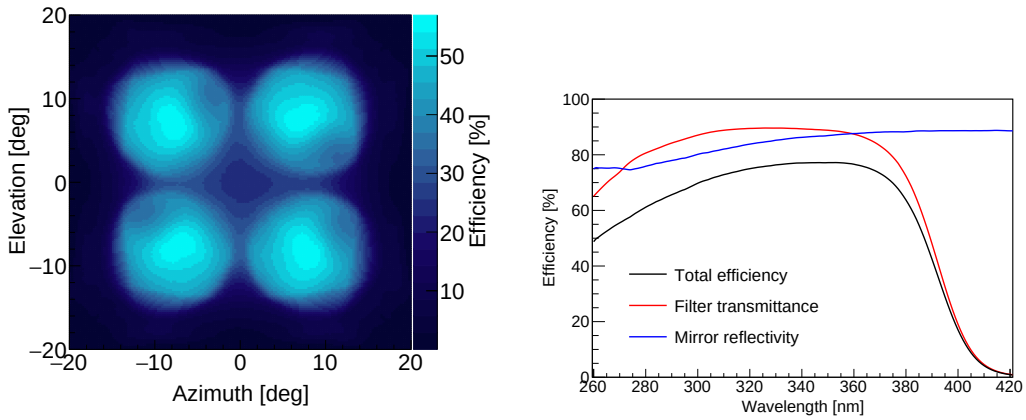
Detailed ray-tracing simulations of the FAST optics were performed with the Zemax software package [223]. The top row of Figure 6.4 shows the results of the ray-tracing simulation of collimated beams. The octagonal aperture of the FAST telescopes causes the "star" shape of the PSF. The simulations of the optical spot are compared with the results of in-situ measurements of the optical PSF which are shown in the bottom row of Figure 6.4. The total directional efficiency of the FAST telescope is shown in Figure 6.3a. The simulation takes into account the full optical system, including the aperture, mirror and camera [217, 223]. There is also an area of $\sim$ 25 % lower efficiency positioned directly opposite the first dynode of each PMT, which arises due to the distance between the photocathode and the first dynode within the large FAST PMTs. This positional inefficiency can be observed as an area of lower efficiency in each PMT in Figure 6.3a. This inefficiency is taken into account in the ray-tracing simulation. The positional dependent relative efficiency was measured in the laboratory at Chiba University [217, 224].

[1] The PSF describes the spatial distribution of light on the focal surface. It is a function of the spherical aberration of the system and is typically circular in shape for both on and off-axis beams.



(a) (b)

Figure 6.2. (a) Design of the full-scale FAST prototype telescope frame, including four PMTs at the focus of a 1.6 m diameter segmented mirror. The support structure is made from aluminium profiles. The UV filter is attached to the outside of the camera box. (b) The dimensions of the optical system of the FAST prototype telescopes. D_a is the diameter of the aperture, D_i is the width and height of the camera box, D_m is the diameter of the segmented mirror, and l is the distance between the mirror and the aperture [217].



(a) (b)

Figure 6.3. (a) Spectrally independent directional efficiency of the FAST camera. The contours show the efficiency of the FAST camera based on the full ray-tracing simulation of the entire FAST optical system [217, 223], and (b) spectral efficiency of the FAST mirror and UV filter as a function of wavelength [217].

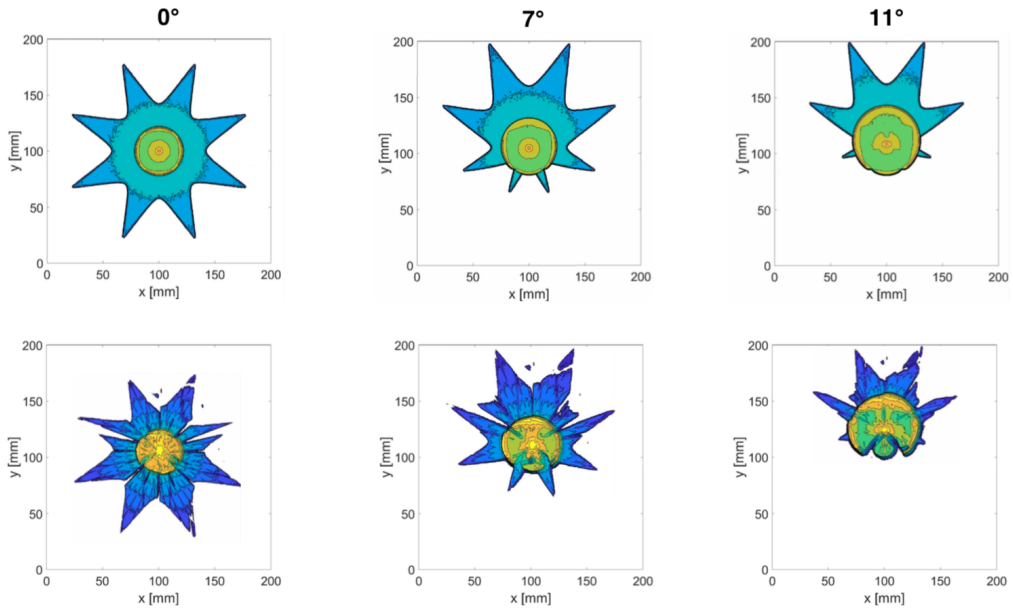


Figure 6.4. The simulated (top panels) and measured (bottom panels) PSF of the FAST prototype telescopes. From left to right are on-axis, 7°, and 11° incidence angles. The true PSF was measured in-situ at the telescope installation site [217].

6.1.3 Data Acquisition and Electronics

The FAST electronics and data acquisition system is designed entirely from commercially available components. The camera consists of four 200 mm Hamamatsu R5912-03 PMTs which are AC-coupled. A CAEN N1470 NIM-mounted module provides the four PMTs with a high voltage. The PMT signals are sent through 15 MHz low-pass filters with a 9^{th} Order Chebyshev response to remove high frequency noise. The signals are subsequently amplified by a factor of 50 using a Phillips Scientific 777 fast amplifier. The resultant signal is digitised at a 50 MHz sampling rate using a 16-channel, 14-bit Struck Innovative Systeme SIS3316 FADC hosted in a portable VME crate along with a Hytec 2092 GPS module to provide event trigger time stamps. Triggers can be provided internally or externally. Internal triggers are implemented as a high-threshold trigger in the data acquisition (DAQ) software. External triggers are supplied via a NIM pulse input to the FADC. Upon triggering, the FAST DAQ system samples a 100 μs data frame from each PMT. A buffer of 10 μs before the trigger time is added in order to estimate the pedestal and baseline variance. The FAST prototypes are operated using a fully remote connection.

The full cost of an individual FAST telescope is ~$25k US. This includes the optical system, mechanical frame, and the data acquisition electronics [217]. A custom FGPA-based data acquisition system with a miniaturised high-voltage supply is currently under development for future iterations of the FAST telescope [217, 221].

6.2 Measurements with FAST Prototypes

6.2.1 Calibration

All FAST PMTs are tested and calibrated in the laboratory at the University of Chicago. The absolute calibration procedure involves installing a PMT in a light-tight test box. The detection (collection plus quantum) efficiency of each PMT is measured to provide the absolute calibration. Yttrium-Aluminium-Perovskite (YAP) pulsers are installed on two of the three prototype telescopes at the TA site. These are used for relative calibration measurements [217]. Each YAP pulser comprises a UV light source produced by the pairing of a Ce-doped $YAIO_3$ scintillator crystal with a 50 Bq ^{241}Am α-source. The YAP pulser generates UV photons at a rate of about 50 Hz with a peak wavelength of 370 nm and a 20 ns FWHM pulse width.

6.2.2 Atmospheric Monitoring

As with any FD telescope, measurements with the FAST telescopes rely on sound estimates of the atmospheric conditions. Typically, the quality of the atmosphere is among the most significant sources of systematic uncertainty for FD measurements, since the atmosphere acts as a calorimeter. It is important to understand conditions such as cloud coverage and atmospheric aerosol content since these can impact how light from air showers is transmitted and scattered through the atmosphere, therefore having a significant effect on the light measurements at the aperture of the FAST telescopes. Several atmospheric monitoring tools are used in conjunction with the FAST prototype telescopes to track the observing conditions at the telescope sites. TA and Auger FD buildings are equipped with weather stations which monitor wind and rain. Both experiments also operate a Central Laser Facility (CLF) which can be used to estimate the atmospheric aerosol content as well as to determine if measurements are likely to be affected by clouds. Data from these instruments have been made available to the FAST collaboration.

At the TA site, the CLF is located at a distance of 21 km from the FAST telescope site. The TA CLF fires 300 vertical laser shots every 30 minutes. These laser shots are observed in the FoV of the central FAST telescope [217]. The path of the CLF laser across the FAST FoV is shown in Figure 6.5a and the average of 200 CLF measurements with the central FAST telescope is shown in Figure 6.5b.

The FAST collaboration has also developed two purpose-built atmospheric monitoring devices; the FAST all sky camera (FASCam) and the sky quality monitor (SQM). The FASCam is designed to measure the cloud coverage above the FAST detectors. It is a

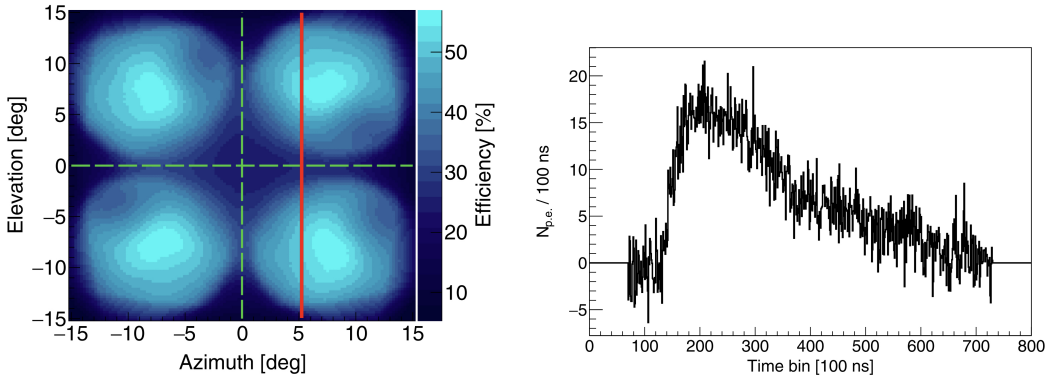


Figure 6.5. (a) The path of the CLF at TA (red) across the FoV of the FAST camera and (b) an example detection of the CLF (an average of 200 laser shots) with the central FAST telescope [217].

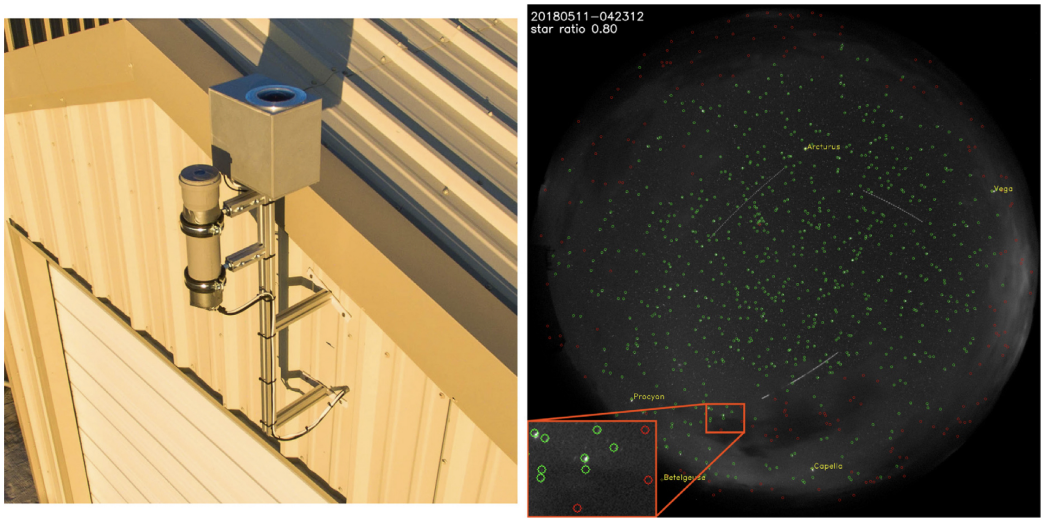


Figure 6.6. The FASCam and SQM (left) and an example of the FASCam analysis (right). Stars are circled in green if they are identified and red if they should be visible but are obscured by cloud [217].

fully automated camera with a FoV of ∼180°. Images of star positions are compared with known coordinates from the Tycho-2 catalogue by an astrometry-based analysis. An example of this analysis is shown in Figure 6.6 (right). The SQM is used to measure the night-sky brightness. A photograph of the FASCam and SQM is shown in Figure 6.6. All data from the FASCam and the SQM are stored in a database, which can be queried during air shower reconstructions [217, 225].

6.2.3 Measurement of Air Showers

Since their installation, the three prototype FAST telescopes have operated in coincidence with the BRM FD at TA. As of March 2019, the prototype telescopes have

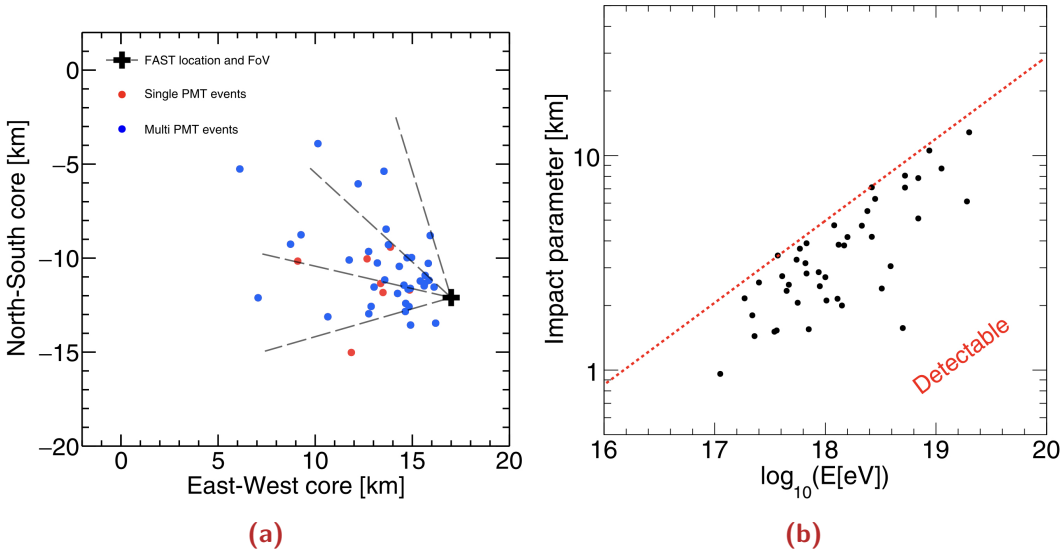


Figure 6.7. (a) The core positions and (b) distance of closest approach as a function of energy for 44 highly significant events detected by FAST in-coincidence with TA. The reconstructed shower parameters are from the Telescope Array (TA) monocular reconstruction [217].

recorded a total of 515 hours of data. A search for showers detected by both the BRM FD and FAST was conducted over $\sim$ 150 hours of data extending to late 2018. The search identified 44 highly significant showers in time coincidence with TA events. They were identified by applying a finite response trapezoidal filter to each trace and rejecting non-shower triggers such as CLF shots [217]. The core locations of these showers with respect to the FAST telescope locations are shown in Figure 6.7a. Additionally, the correlation between the distance of closest approach (impact parameter) and the energy of the events is shown in Figure 6.7b. It is possible to approximately estimate the spacing for the FAST stations based on the sensitivity of the prototype telescopes. The red line in Figure 6.7b indicates the maximum detectable impact parameter for showers of a given energy. Extrapolating based on the events observed so far suggests a spacing of $\sim$20 km at the FAST target energy of $10^{19.5}$ eV.

The highest energy event in the sample was observed by the FAST prototypes on the 15$^{\text{th}}$ of May 2018. The PMT signals measured by FAST 1 and 2 are shown in Figure 6.8. According to the TA monocular reconstruction, the event has an energy of $\sim$19 EeV and a zenith angle of $\sim$55°. This event is dominated by Cherenkov light [217]. An example of a typical fluorescence dominated event is shown in Figure 6.9. This event was observed by the FAST prototypes on the 11$^{\text{th}}$ of May 2018, with the TA monocular reconstruction providing an energy of $\sim$4 EeV and a zenith angle of $\sim$57° [217].

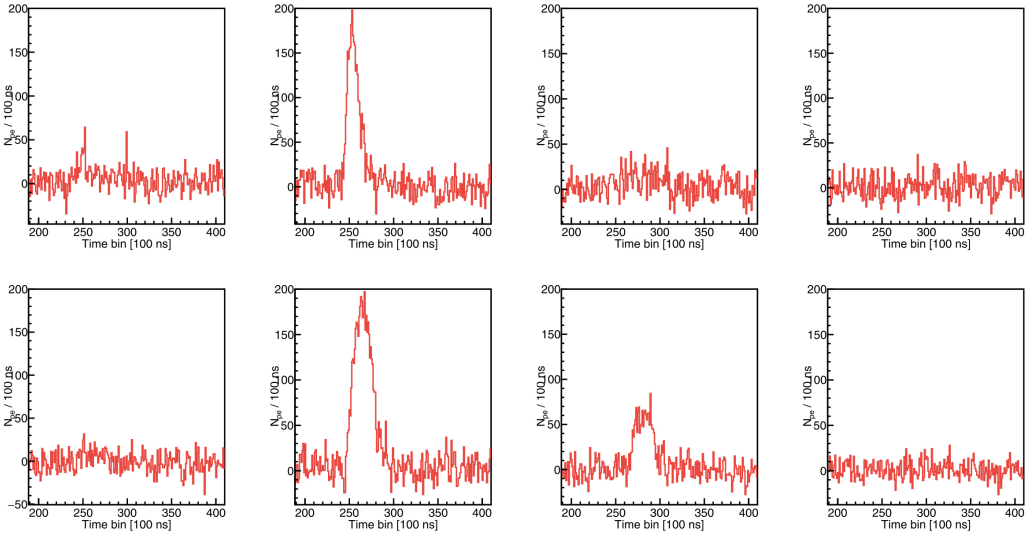


Figure 6.8. The FAST PMT traces from the highest energy event measured on the 15th of May 2018. The energy of the event is ∼19 EeV and its zenith angle is ∼55°. The eight traces are arranged based on the sky view of the PMTs [217].

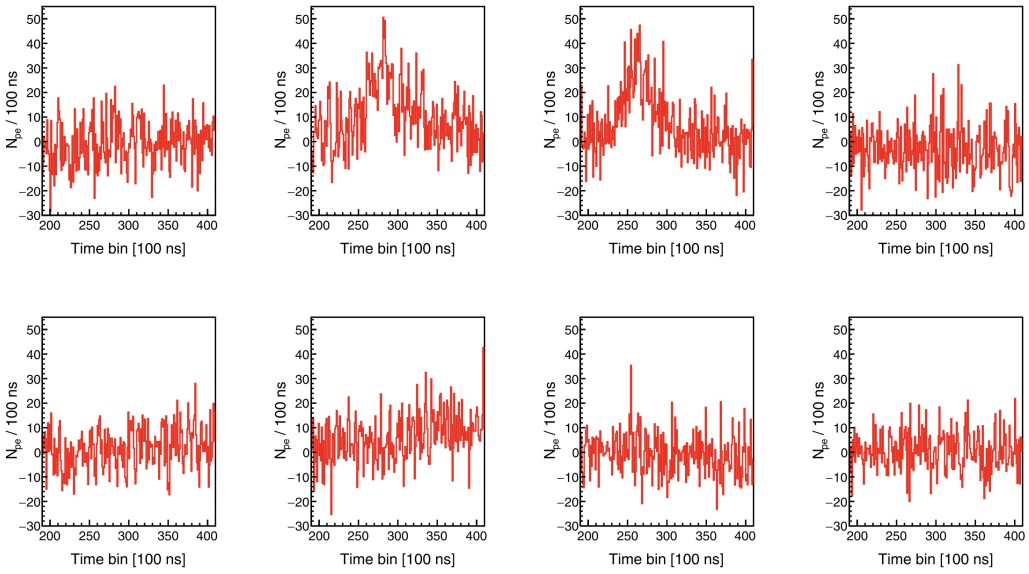


Figure 6.9. The FAST PMT traces from a fluorescence dominated event measured on the 11th of May 2018. The energy of the event is ∼4 EeV and its zenith angle is ∼57°. The eight traces are arranged based on the sky view of the PMTs [217].

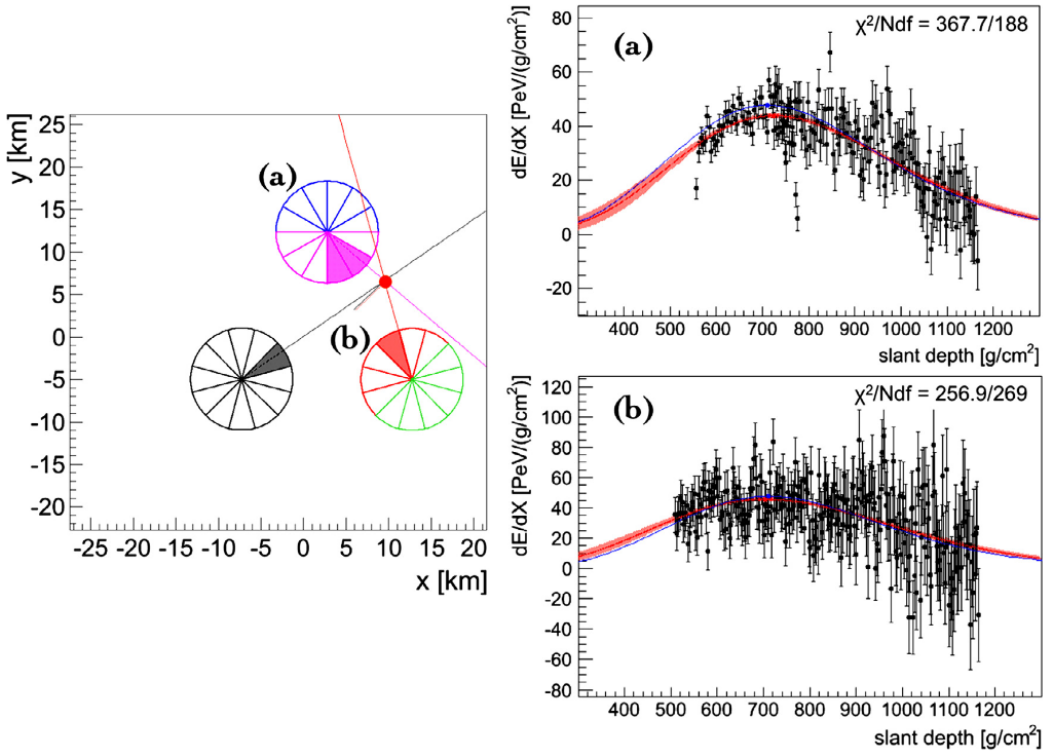


Figure 6.10. An example reconstruction of a simulated shower with energy $10^{19.5}$ eV. The shower geometry is the smeared MC truth. Left: the layout of the simulated FAST stations including their FoV. The shower core is indicated by the red point. Right: the reconstructed (red) energy deposit profiles for the simulated shower (blue) for the two closest stations labelled (a) and (b) [215].

6.3 Expected Performance

Due to the FAST camera's low resolution, standard geometry reconstruction like that of the Auger FD reconstruction is not possible. Reconstruction of the shower-detector plane (SDP) relies on the identification of a line of triggered pixels across the Auger FD camera; something that is clearly impossible for the four-pixel FAST camera. For high energy showers which trigger multiple FAST stations simultaneously, the timing information between the stations can be used to constrain the geometry. However, the use of a ground array similar to the Auger SD may be required to provide adequate geometry reconstruction. This concept was investigated using extensive simulations of the FAST design in [215]. In order to emulate the effect of a SD array, the true geometries of the simulated showers were smeared by 1.0° in arrival direction and 100 m in core location (the typical resolution of existing SD arrays like Auger [136]). A triangular arrangement of FAST stations with a spacing of 20 km was used to detect and reconstruct simulated air showers generated by CORSIKA. This was done using a modified version of the Auger Offline software. Figure 6.10 shows an example reconstruction of such a simulated event. Given the geometry, the energy deposit profile can be reconstructed by unfolding the detector efficiency

and atmospheric attenuation. A Gaisser-Hillas profile fit was fit to the shower profile in order to determine the total energy and depth of maximum of the shower. Using selected events with X_{max} in the FoV and additional quality cuts, analysis of the reconstruction efficiency, energy and X_{max} resolutions of the simulated FAST detector were performed. The results of this study are shown in Figure 6.11. As expected, FAST performs better at energies above $10^{19.5}$ eV, where the reconstruction efficiency is 100%, energy resolution is $\sim$10% and X_{max} resolution is $\sim$34 g/cm^2. This is comparable to current generation FD reconstruction [215]. Improved techniques of event reconstruction (especially development of a geometrical reconstruction) using the FAST telescopes is one of the main investigations of this thesis.

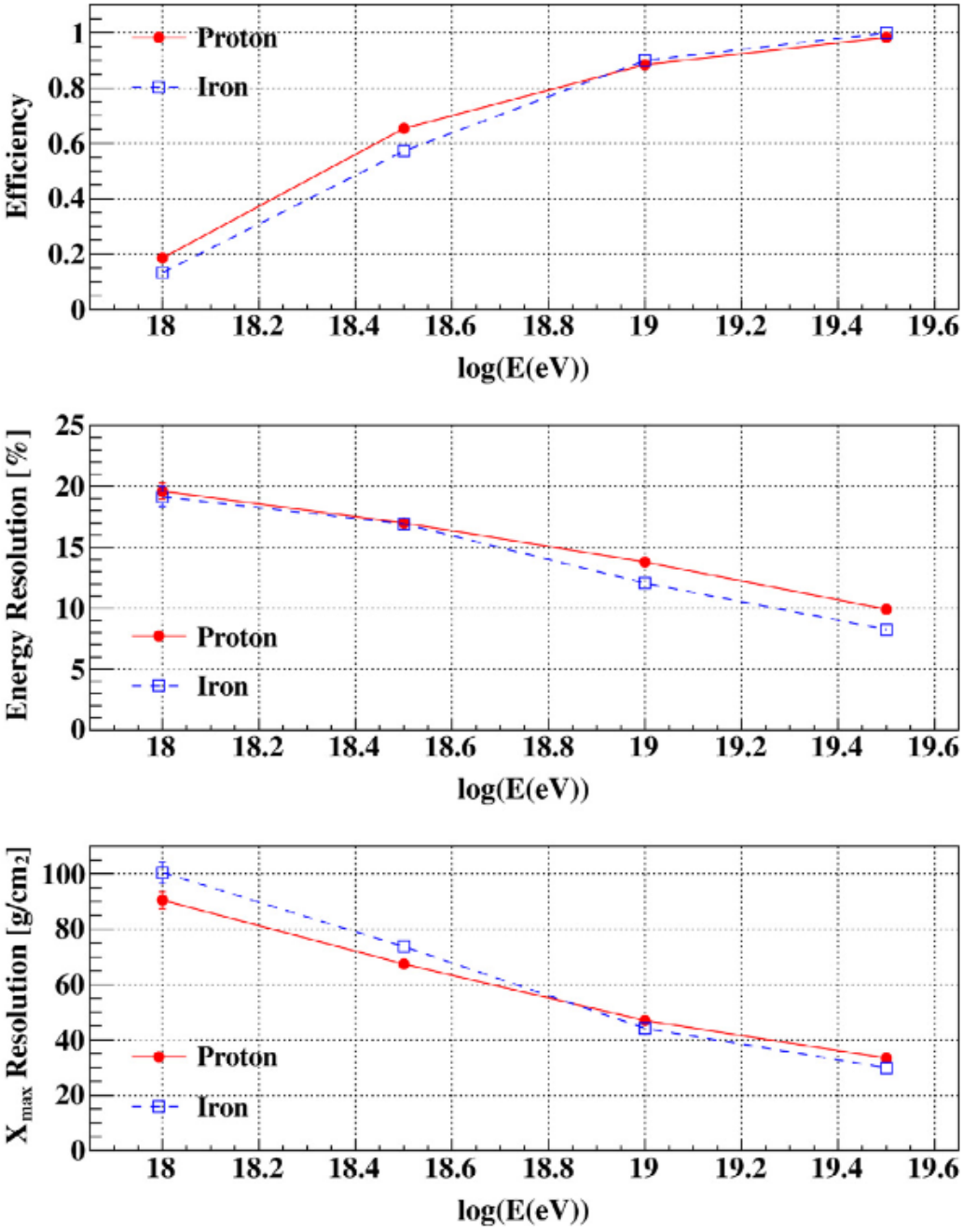


Figure 6.11. From top to bottom: FAST reconstruction efficiency, energy resolution and $X_{\max}$ resolution as functions of energy shown for proton (red) and iron (blue) simulated showers [215].

7 Top-Down Reconstruction

7.1 Introduction

In order for the Fluorescence detector Array of Single-pixel Telescopes (FAST) to measure the cosmic ray energy spectrum and mass composition of the highest energy cosmic rays, it is vital to be able to reconstruct extensive air showers (EAS) as accurately and completely as possible. In Chapter 6 it was established that the significantly reduced resolution of FAST means that it is impossible to use standard techniques to reconstruct important shower parameters from the measured signals. Traditional air shower reconstruction techniques use a bottom-up approach in which only a subset of the available information is utilised to fit the shower parameters. These techniques generally follow a two-step approach: 1) fit for the shower geometry (zenith, azimuth and core position) using a track of triggered pixels across the camera, and 2) fit a Gaisser-Hillas profile to the energy deposited as a function of slant depth. In the first step of the geometry fit, only two parameters are extracted from each photomultiplier tube (PMT) trace in order to fit the shower geometry, namely, the total integrated signal and the centroid time (signal-weighted time average) [136]. This technique works well for a finely pixelated camera where the total signal and centroid time of each pixel, along with its individual pointing direction is enough to constrain the shower axis within an acceptable resolution. This is depicted in Figure 7.1.

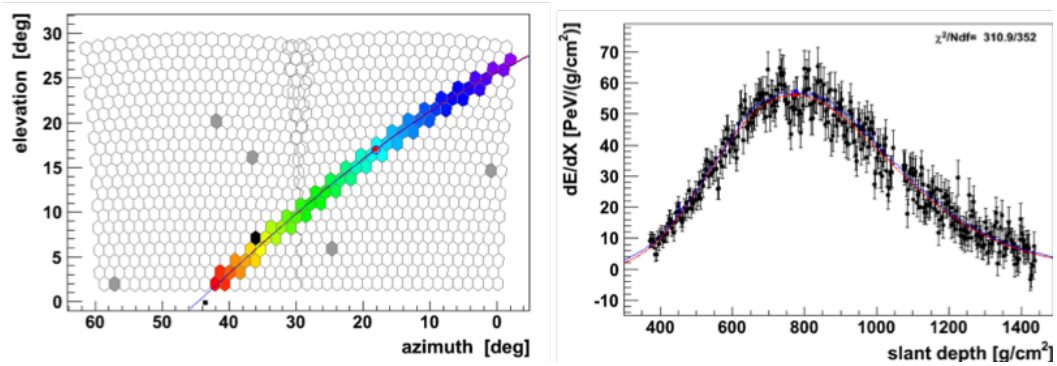


Figure 7.1. Left: Track of triggered pixels across the camera of the Pierre Auger Observatory (Auger) Fluorescence Detector (FD). Right: The Gaisser-Hillas profile fit to the energy deposited as a function of atmospheric depth.

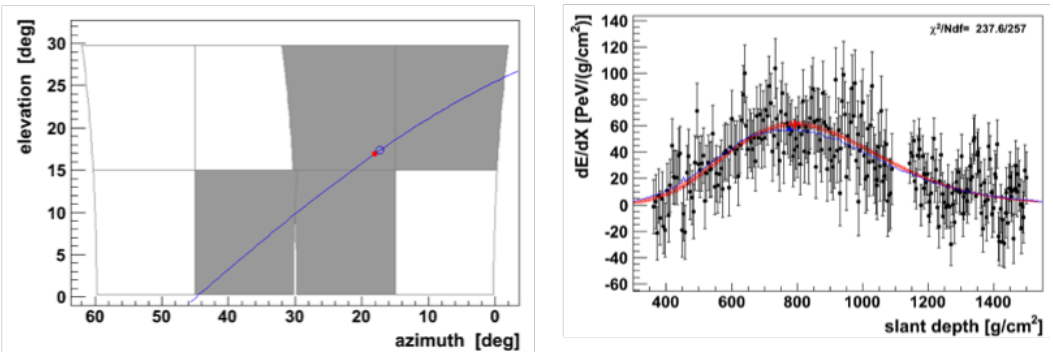


Figure 7.2. Left: The simulated shower axis (blue line) across the four pixels of the FAST camera. Right: The Gaisser-Hillas profile fit to the energy deposited as a function of atmospheric depth from a simulation of the FAST detector implemented in the Auger Offline framework.

Since the FAST design contains only four pixels over the same field of view (FoV) as standard Auger FD telescopes, as shown in Figure 7.2, there is a requirement for a more sophisticated reconstruction technique in order to determine shower parameters with acceptable resolution. Furthermore, without a suitable geometry reconstruction, the shower profile cannot be determined accurately. The top-down reconstruction technique outlined in this chapter addresses the requirement for a full FAST reconstruction including a reconstruction of the shower geometry. Alternatively, the top-down reconstruction can be used in conjunction with the shower geometry as determined by a Surface Detector (SD). By taking into account the full information available from the measurements at the level of individual PMTs, the technique will provide the best possible accuracy for a FAST reconstruction.

The top-down algorithm attempts to derive information about air shower parameters using a maximum likelihood technique which directly compares an event's measured PMT traces to the PMT traces of simulated events. Based on a good understanding of the detector and its response, photons from simulated air showers are propagated through a realistic simulation of the FAST detector to produce simulated traces. Using first guess estimates of the shower parameters, along with the maximum likelihood function and minimisation algorithm implemented with Minuit, conclusions about the agreement between simulated and measured showers can be drawn and the most probable values of the shower parameters, including their uncertainties, can be estimated. The reconstruction algorithm presented in this chapter is a new approach to a similar algorithm developed by Schüssler [226] for Auger. The technique was shown then to produce significant improvements in the resolution of reconstructed parameters compared to the standard reconstruction procedure. However, due to the highly computational nature of the top-down reconstruction procedure, it was proposed to be used only for rare ultra-high energy events. For FAST, however, such improvements in the resolution are essential given the limitations of its four-pixel design.

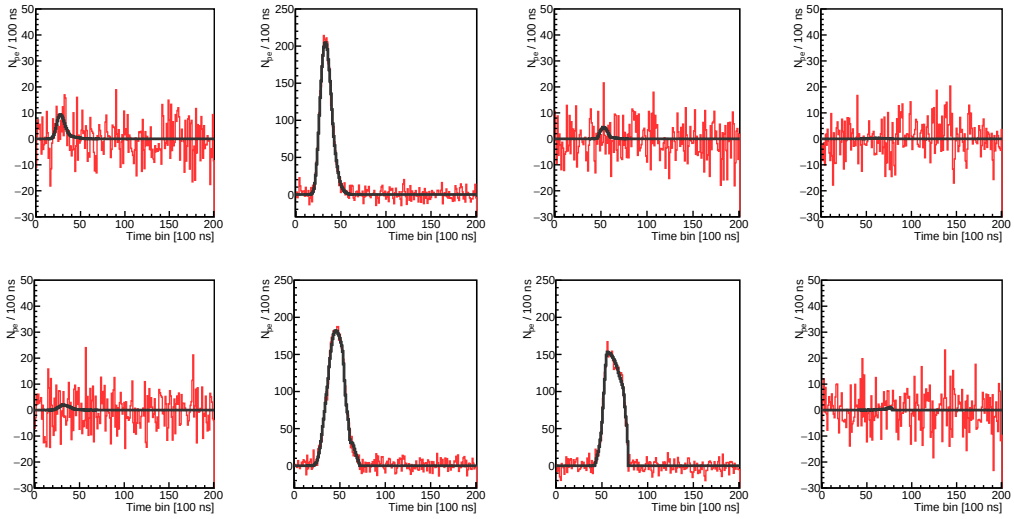


Figure 7.3. Example of a comparison between a simulated event with (red) and without (black) added background noise. The eight PMT traces are arranged in a sky-view orientation. The shower is a simulated version of the highest energy event measured by the FAST prototype telescopes at the Telescope Array (TA) presented in Section 6.2.3. It has an energy of 19 EeV and a zenith angle of 55°.

7.2 Reconstruction Technique

In order to reconstruct shower parameters with the FAST telescopes, a maximum likelihood comparison between measured data and a semi-analytical simulation has been developed. The reconstruction algorithm uses all available information at the level of individual PMT traces in order to simultaneously reconstruct the shower geometry, X_{max} and energy. An example comparison of the PMT traces between a simulated shower with added noise, and without added noise is shown in Figure 7.3.

7.2.1 Likelihood Function

The reconstruction of air showers is a specific case of the more generalised problem of estimating a set of unknown parameters $\vec{a}$ given a set of measured values $\vec{x}$. In this case, the parameters $\vec{a}$ are the shower geometry (θ, ϕ, x, y), atmospheric depth of shower maximum, X_{max} and energy. The most probable values of these parameters, along with their uncertainties, are determined by maximizing the likelihood function, $\mathcal{L}(\vec{x}\,|\vec{a}\,)$, which for N independent measurements x_i is the product of the probabilities for measuring each x_i given parameters $\vec{a}$

$$\mathcal{L}(\vec{x}\,|\vec{a}\,) = \prod_{i}^{N} P(x_i|\vec{a}\,) \tag{7.1}$$

The likelihood function is maximised when the predicted (simulated) signals best match the measured signals. In practice, the log-likelihood is maximised instead since the log-likelihood is given by the sum of the probabilities

$$\ln \mathcal{L}(\vec{x}\,|\vec{a}\,) = \sum_{i}^{N} P(x_i|\vec{a}\,). \tag{7.2}$$

For convenience and consistency with χ^2 fitting, the reconstruction is processed by minimising $-2\ln\mathcal{L}$ with respect to the shower parameters $\vec{a}$. Due to the complexity of the likelihood function and its simulation requirements, it is not possible to minimise $-2\ln\mathcal{L}$ analytically. Instead, numerical methods must be employed in the minimisation. One option is to generate a pre-simulated library of events on a grid which spans the full parameter space with step sizes smaller than the resolution of the shower parameters. Alternatively, the simulation can be built into the reconstruction procedure where the minimisation algorithm can be used to more efficiently search the parameter space for the best match. The additional benefits of simulating events during the minimisation procedure is that time dependent characteristics of the detector such as calibration, operation, and even atmosphere can all be accounted for by the reconstruction by simply modifying the configuration of the detector simulation. Therefore, the approach of simulating events during the minimisation procedure is taken for the top-down reconstruction. The disadvantage of this approach is the computationally expensive nature of running many detector simulations in order to reconstruct each event. However, the advantage of being able to easily account for time dependent detector effects outweighs the computational disadvantages.

7.2.2 Signal Uncertainty

The probability of observing a signal within an individual time bin is dependent on the expected signal and background fluctuations. The dominant source of background for FAST is the night-sky background (NSB) [217]. Since the detector electronics are AC-coupled, the background light cannot be measured directly. Instead, the variance of the noise in early time bins of the measured trace (which are free from any shower signal) can be used as an estimate for the total background noise measured by a PMT. This can be done based on the assumption of a Poisson distributed background. For a large number of background photons, the Poisson assumption can be relaxed to a Gaussian distribution with mean and variance both equal to the Poisson mean. Since the sum of independent, random normal variates is also distributed normally, no special consideration needs to be made for independent sources of background light, for example, from bright stars or the elevation dependence of the NSB. Additional fluctuations are due to electronic noise and the

inherent gain variance of the PMTs. The total background σ^2_{bkgd} (measured in units of photoelectrons) within a single time bin is given by

$$\sigma^2_{\text{bkgd}} = n_{\text{bkgd}}\,(1 + V_g) + \sigma^2_{\text{elec}} \tag{7.3}$$

where n_{bkgd} is the mean number of photoelectrons measured from the NSB, V_g is the gain variance of the PMT and σ^2_{elec} is the electronic noise. In practice, the combined background variance, σ^2_{bkgd} is measured rather than the individual components of NSB and electronic noise.

Since signal photoelectrons are also Poisson distributed, the total variance σ^2 is given by

$$\sigma^2 = \left(n_{\text{bkgd}} + n_{\text{signal}}\right)(1 + V_g) + \sigma^2_{\text{elec}} \tag{7.4}$$

where n_{signal} is the number of photoelectrons from shower signal. For a large number of photoelectrons, this is well represented by a Gaussian with a variance of $\sigma^2_{\text{bkgd}} + n_{\text{signal}}(1 + V_g)$.

7.2.3 Signal Probability Density Function

Based on the background model discussed in Section 7.2.2, the probability of observing a signal given the background noise can be derived. The probability of observing a signal, x based on an expected signal μ with variance $\sigma^2 + \mu(1 + V_g)$ is given by

$$P(x|\mu, \sigma, V_g) = \frac{1}{\sqrt{2\pi\left[\sigma^2 + \mu\,(1 + V_g)\right]}} \exp\left[-\frac{(x - \mu)^2}{2\left(\sigma^2 + \mu\,(1 + V_g)\right)}\right] \tag{7.5}$$

Hence, the full probability density function for a single pixel with a trace of N time bins is the product of the probabilities for each individual time bin

$$P(x|\mu_i, \sigma, V_g) = \prod_i^N \frac{1}{\sqrt{2\pi\left[\sigma^2 + \mu_i\,(1 + V_g)\right]}} \exp\left[-\frac{(x_i - \mu_i)^2}{2\left(\sigma^2 + \mu_i\,(1 + V_g)\right)}\right] \tag{7.6}$$

Using Equations (7.1) and (7.2) to convert to a log-likelihood function for convenience this becomes

$$-\,2\ln\mathcal{L}(\vec{x}\,|\vec{a}\,) = \sum_i^N \frac{(x_i - \mu_i)^2}{\left[\sigma^2 + \mu_i\,(1 + V_g)\right]} + \ln\left[\sigma^2 + \mu_i\,(1 + V_g)\right] + \ln 2\pi \tag{7.7}$$

7.2.4 Energy Scale

The reconstruction of the shower energy can be determined during the reconstruction procedure as one of the six dimensions of minimisation. However, the expected signal μ_i can instead be modified by an energy scale factor, A which is a free parameter controlling the energy fit. This can be implemented by modifying Equation (7.6) to provide

$$P(x|\mu_i,\sigma,V_g) = \prod_i^N \frac{1}{\sqrt{2\pi\left(\sigma^2 + A\mu_i\left(1+V_g\right)\right)}} \exp\left(-\frac{\left(x_i - A\mu_i\right)^2}{2\left(\sigma^2 + A\mu_i\left(1+V_g\right)\right)}\right) \quad (7.8)$$

This is possible since the calorimetric shower energy simply scales the expected signal up or down, so the simulation of different values of shower energy is not necessary. This is beneficial for the timely convergence of the minimisation procedure. In this case, μ_i is the expected signal for an event with a reference energy of 10^{19} eV.

7.2.5 Event Likelihood Function

The log-likelihood can now be defined at the event level given the combined probabilities of measuring signals in each time bin and pixel contained in a given event. The total event log-likelihood function is given as the sum of probabilities over all pixels and signal bins

$$\ln\mathcal{L}(\vec{x}\,|\vec{a}\,) = \sum_k^{N_{\text{pix}}}\sum_i^{N_{\text{bins}}} P_k(x_i|\vec{a}\,) \quad (7.9)$$

where $P_k(x_i|\vec{a}\,)$ is the probability of measuring a signal of x_i photoelectrons in the ith time bin of pixel k.

The likelihood function in Equation (7.9) is maximized with respect to the shower parameters using the ROOT implementation of Minuit2 in C++ . The reconstructed shower parameters $\vec{a}$ are those which maximize the likelihood (or equivalently minimize $-2\ln\mathcal{L}$). Parameter uncertainties are also derived from the 1σ likelihood contours.

7.2.6 Parameter Uncertainties

Given the log-likelihood as a function of the shower parameters, it is possible to determine statistical uncertainties based on the 1σ likelihood contours. That is, the uncertainty in a given parameter is defined by the change in the parameter which results in a 1σ increase in the likelihood function. The value for which the likelihood function must increase by in order to change a parameter by 1σ depends

n	1	2	3	4	5	6	7	8
x_c	1	2.28	3.53	4.72	5.89	7.04	8.18	9.30

Table 7.1. Critical value x_c to increase the value of $-2\ln\mathcal{L}$ by in order to derive parameter uncertainties for n degrees of freedom.

on the number of parameters (degrees of freedom) which are being reconstructed simultaneously. The log-likelihood is closely related to the χ^2 distribution. Note that the first term of Equation (7.7) is identical to the χ^2. The χ^2 distribution is given by

$$P(x,n) = \frac{\left(\frac{x}{2}\right)^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2\Gamma\left(\frac{n}{2}\right)} \tag{7.10}$$

where n is the number of degrees of freedom in the variable x, and Γ is the gamma function. For example, in the case of a two-parameter reconstruction, there are two degrees of freedom (e.g. X_{max} and energy only), so $n = 2$. In this case the χ^2 distribution is given by

$$P(x,1) = \frac{\left(\frac{x}{2}\right)^{1-1} e^{-\frac{x}{2}}}{2\Gamma\left(1\right)} = \frac{1}{2}e^{-\frac{x}{2}}. \tag{7.11}$$

To determine the critical value x_c which corresponds to the 68% confidence level, the integral of the χ^2 probability must be equal to 68% since 1σ contains 68% of the parameter values. The integral of Equation (7.11) is given by

$$P(x < x_c) = \frac{1}{2}\int_0^{x_c} e^{-\frac{x}{2}}dx = 1 - e^{-\frac{x_c}{2}} = 0.68. \tag{7.12}$$

Therefore, the critical value is given by

$$x_c = -2\ln(0.32) \sim 2.28 \tag{7.13}$$

This calculation of x_c can be repeated for any number of degrees of freedom (parameters) n. Table 7.1 shows the approximate values of x_c for a selection of n. All subsequent parameter uncertainties for the top-down reconstruction are determined by the change in each parameter required to increase $-2\ln\mathcal{L}$ by the critical value appropriate for the number of parameters being reconstructed.

7.3 Simulation of the FAST telescope

As discussed in Section 7.2, the FAST reconstruction requires a realistic simulation of the detector response to enable a maximum likelihood comparison at the level of individual PMT traces to reconstruct the shower parameters. The simulation software, `FAST-sim` [217] is implemented in C++ using a modified version of the

Auger Offline software framework [163]. All simulation modules are purpose built for FAST. An example sequence file for the simulation of a single FAST event is shown in Code 7.1.

```
<!-- A sequence for FAST shower simulation -->
<sequenceFile>
  <enableTiming/>
  <moduleControl>

    <loop numTimes="1">

      <module> FASTProfileSimulatorCG    </module>
      <module> FASTEventGeneratorCG      </module>
      <module> ShowerLightSimulatorCG    </module>
      <module> FASTSimulatorCG           </module>
      <module> FASTEventFileExporterUA   </module>

    </loop>

   </moduleControl>
</sequenceFile>
```

Code 7.1 An example sequence file for the simulation of a single FAST event.

The simulation takes the shower parameters as input and returns a full simulation of the individual FAST PMT traces. The approach of the FAST detector simulation is similar to that of the Auger shower profile reconstruction discussed in Section 3.5.2.6, except that the process is reversed in order to calculate the light emitted in the atmosphere along the shower axis given the energy deposited as a function of atmospheric depth from the input shower profile.

The first module, `FASTProfileSimulator` produces an analytical Gaisser-Hillas profile [175] based on the input shower parameters, and subsequently calculates the energy deposited in bins of atmospheric depth. The functional form of the Gaisser-Hillas function is defined as

$$f_{GH}(X) = \left(\frac{dE}{dX}\right)_{X_{\max}} \left(\frac{X - X_0}{X_{\max} - X_0}\right)^{\frac{X_{\max} - X_0}{\lambda}} e^{-\frac{X - X_{\max}}{\lambda}} \tag{7.14}$$

where $(dE/dX)_{X_{\max}}$ is the energy deposited at shower maximum $X_{\max}$, and X_0 and λ are shape parameters. The next module, `FASTEventGenerator` configures the shower timing, geometry and coordinate system-related parameters including the input core position of the simulated shower.

Using the `ShowerLightSimulator` module, the number of fluorescence photons at the shower track is calculated using the AIRFLY fluorescence model, taking into account a realistic parametrisation of a desert atmosphere. The Cherenkov light

contribution is also calculated based on the number of shower electrons above the Cherenkov threshold in air [217].

The FAST detector simulation procedure is handled by a single module, `FASTSimulator`. This includes propagating photons through a parametrised molecular and aerosol atmosphere to each FAST telescope, taking into account the wavelength-dependent atmospheric transmission. The simulation takes into account direct fluorescence and Cherenkov photons, as well as Cherenkov photons scattered out of the beam along the shower axis. Photons are then propagated through the FAST optics using the measured response of the ultra-violet (UV) filter and mirror reflectivity (Figure 6.3b). The signals are weighted by the full ray-tracing simulation of the FAST camera, accounting for the directional sensitivity, detection efficiency and optical spot size (Figure 6.3a) discussed in Section 6.1.2, as well as the PMT quantum efficiency. The signal in each PMT is calculated in units of photoelectrons per 100 ns in order to generate the PMT traces. Finally, the results of the simulation are written into the `FASTEventFile` format using the `FASTEventFileExporter` module.

7.4 Reconstruction Algorithm

The reconstruction software `FAST-rec` is implemented in C++ using the FAST modified version of the Auger Offline software framework described in Section 7.3. The author has developed all reconstruction software, including a new event structure which utilises the `ROOT` file format. A description of this FAST event structure is provided in Appendix B to compliment the work presented in this chapter. An example sequence file for the reconstruction of FAST events is shown in Code 7.2.

```
<!-- A sequence for FAST shower reconstruction -->
<sequenceFile>
  <enableTiming/>
  <moduleControl>

    <loop numTimes="unbounded" pushEventToStack="yes">

      <module> FASTEventFileReaderUA       </module>
      <module> FASTTopDownReconstructorUA  </module>
      <module> FASTEventFileExporterUA     </module>

    </loop>

   </moduleControl>
</sequenceFile>
```

Code 7.2 An example sequence file for the reconstruction of FAST events.

The first module, `FASTEventFileReader` processes a `FASTEventFile` and extracts the `FASTEvent` from the input ROOT file. The input event can either be a simulated event generated from the `FAST-sim` program (Section 7.3) or a real measured event. The second module, `FASTTopDownReconstructor` is the main processing module of the reconstruction program. This module handles the minimisation procedure as well as the processing of the simulation modules at each step of the minimisation. This involves running the first four modules shown in Code 7.1 based on the requested shower parameters from Minuit. The final module, `FASTEventFileExporter` is identical to that used in `FAST-sim` described in Section 7.3. The steps in the reconstruction algorithm are explained in detail below.

7.4.1 Pixel Calibration

The first step of the reconstruction procedure is the conversion from ADC counts to photoelectrons in order to compare with the output of the detector simulation. As discussed in Chapter 6, this is done using the calibration constants obtained from laboratory measurements and on-site relative calibrations.

Next, the baseline is subtracted from the trace of each pixel. As discussed in Section 7.2, the FAST camera is illuminated by significant NSB light which cannot be measured directly since the FAST electronics are AC-coupled. This form of noise dominates all other background sources including electronic noise. However, fluctuations around the baseline are a combination of both the NSB and electronic noise. These fluctuations must be determined as a part of the reconstruction procedure. The baseline and its variance are estimated from the signals in early time bins (bin 1-200) of the measured trace. These bins are free from any shower signal and therefore contain purely background signal.

7.4.2 Likelihood Search

Once the measured traces have been calibrated, they can be compared with simulated traces. During the minimisation procedure, the measured trace is compared to simulated traces from many variations of the shower parameters. The traces from simulated events with shower parameters very similar to that of the measured event will match more closely than those with very different shower parameters. This is illustrated in Figure 7.4 where the traces from a single telescope from a simulated shower with an X_{max} of 800 g/cm^2 and an energy of $10^{19.5}$ eV are compared to simulated traces with (a) identical shower parameters, and (b) an X_{max} of 900 g/cm^2 with all other shower parameters fixed to their generated values. In the first case where all shower parameters are identical, the match is perfect. However, in the

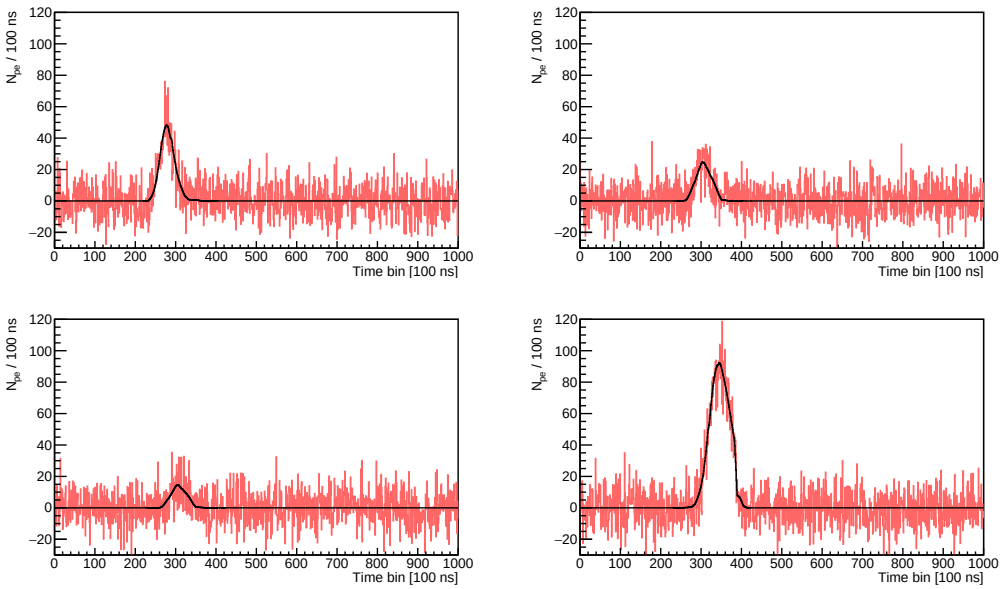
case where X_{max} was shifted by 100 g/cm^2, visible differences between the traces arise. While the differences are small in magnitude, they are statistically significant, especially when coupled with similar differences in the traces of several additional telescopes which may have measured the event in stereo mode for high energy events.

For each iteration of the minimisation procedure, the measured traces from all telescopes are compared to the simulated traces of the test event and the value of $-2\ln\mathcal{L}$ is calculated from Equation (7.9). The shower parameters are varied in order to achieve a smaller value of $-2\ln\mathcal{L}$. The trace comparison and likelihood calculation are repeated for each simulated trial event. The minimisation procedure continues until the algorithm converges and a minimum has been determined. The algorithm attempts to find the shower parameters (X_{max}, energy, θ, ϕ, x and y) which provide simulated PMT trace signals which best match the measured traces. The shower parameters which minimise $-2\ln\mathcal{L}$ are taken as the best estimate of the reconstructed shower parameters.

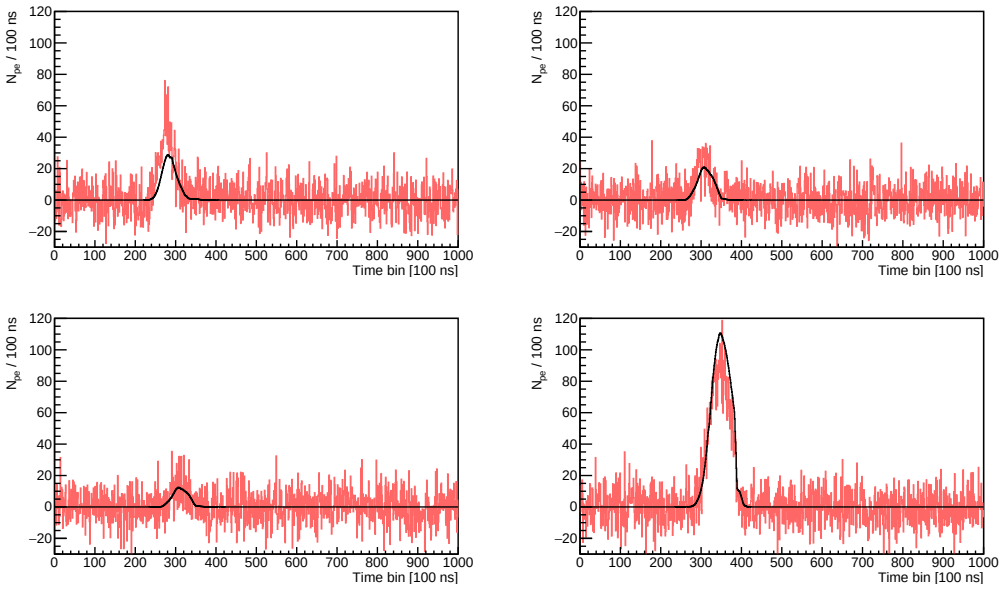
7.4.3 Absolute Time Offset

During the reconstruction process, the absolute time offset between the data and simulated traces must be determined. The timing (and trigger) of each event is dependent on background fluctuations. However, the absolute time of an event itself has no physical meaning. Only the relative timing of the pulses between pixels provides meaningful information about the shower parameters, particularly the shower geometry. While maintaining the relative timing information between PMTs, the absolute time of all traces in an event are shifted to find the best-fit absolute time offset between the data and simulated traces. While the actual value of the time offset is not important, it is important that the most suitable comparison is made between the data and simulated traces. A misalignment in the measured and simulated traces vastly changes the value of the likelihood determined using Equation (7.9).

During the reconstruction procedure, the best-fit absolute time offset is determined for each iteration of the minimisation. For a given simulated test event, the measured traces are all shifted in steps and the value of $-2\ln\mathcal{L}$ is recalculated using Equation (7.9) given the new absolute time offset. For the given simulated event, the absolute time offset which minimises the value of $-2\ln\mathcal{L}$ is used to align the measured and simulated traces. Once the absolute time offset has been determined, the value of $-2\ln\mathcal{L}$ is passed back to the minimiser to continue with the next iteration of the minimisation of the shower parameters. An example reconstruction of the absolute time offset using this procedure is shown in Figure 7.5 where an



(a)



(b)

Figure 7.4. Example of PMT traces (red) from an event compared with (a) the noise-free traces from an event with the generated shower parameters and (b) with $X_{\max}$ shifted by 100 g/cm^2 with all other parameters fixed. The top-left and bottom-right PMTs show noticeable differences in their traces with $X_{\max}$ shifted by 100 g/cm^2. The four PMT traces are arranged based on the sky view of the simulated FAST telescopes.

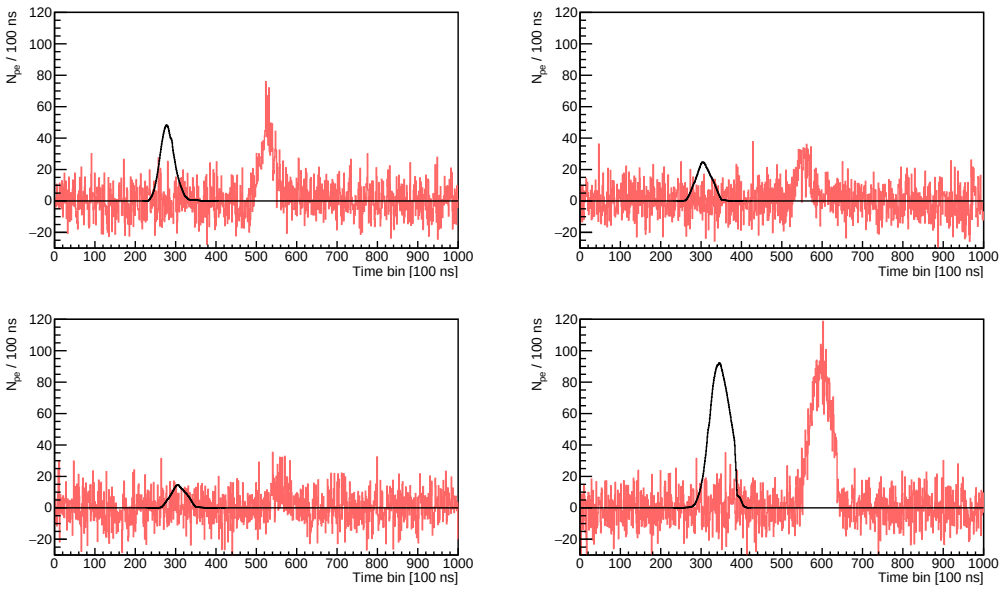
artificial absolute time offset of 25 μs (250 bins) was applied to all traces. All shower parameters have been fixed in this example. The initial comparison in Figure 7.5a clearly shows the misalignment of the traces. The final comparison after the reconstruction procedure has been completed is shown in Figure 7.5b where the absolute time offset was correctly determined to be 25 μs (250 bins). The value of $-2\ln\mathcal{L}$ as a function of the time shift is shown in Figure 7.6 for this test.

7.5 Testing the Reconstruction Algorithm on Simulated Events

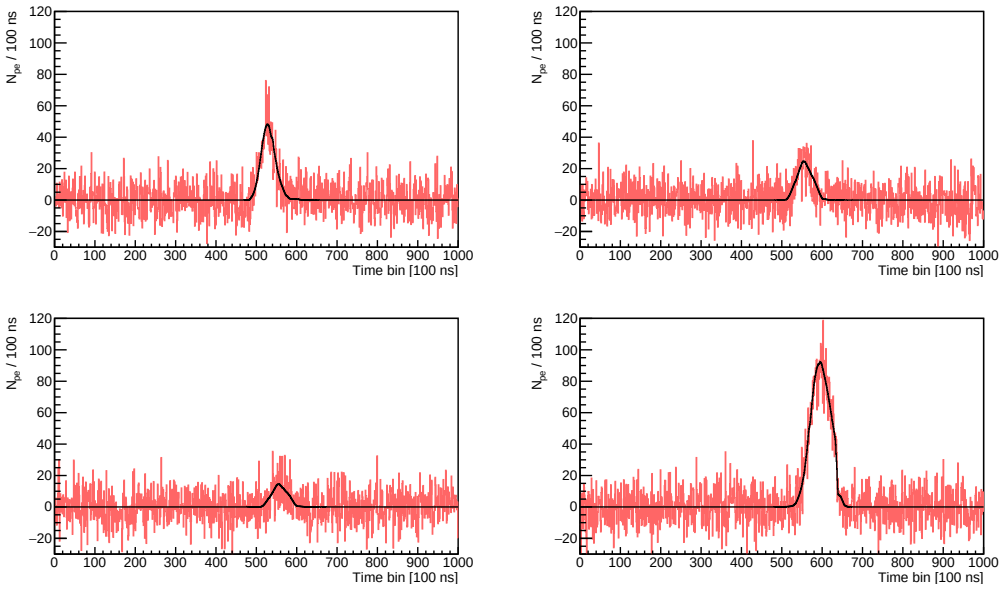
Before a full FAST reconstruction is attempted, it is important to test the reconstruction algorithm with simple examples to understand the strengths and weaknesses of the simulation and reconstruction technique. In this section the reconstruction algorithm is tested on sets of simulated events generated using the FAST simulation described in Section 7.3 in order to validate the top-down reconstruction. For the purpose of these tests, a dataset containing 1000 events has been simulated to study the reconstruction performance in the simple case where only a single parameter is reconstructed. The events are simulated at a fixed energy of $10^{19.5}$ eV, X_{max} is sampled from a parametrisation of the EPOS-LHC hadronic model, the arrival directions are sampled from a realistic $\sin\theta\cos\theta$ distribution, and the core positions are sampled uniformly within a 10 km radius of the centre of a triangular cell of FAST stations as shown by the dashed circle in Figure 7.7. During reconstruction, all shower parameters are fixed to their generated values except the parameter of interest. While this is not practical for measured data, it serves the important purpose of testing the top-down reconstruction for a very simple case without the need for complicated multi-dimensional visualisation of the shower parameters.

7.5.1 X_{max} Reconstruction

The value of $-2\ln\mathcal{L}$ as a function of X_{max} for an example event is shown in Figure 7.9. A clear minimum over a large range of X_{max} is visible. For this particular simulated event, the reconstructed value of X_{max} is 851.78 ± 3.04 g/cm^2, which contains the true X_{max} of 850 g/cm^2 within statistical uncertainty. The reconstruction is repeated for each of the 1000 simulated trial events. The reconstructed X_{max} and corresponding 1σ uncertainty is obtained from Minuit. A distribution of the difference between the simulated and reconstructed X_{max} is shown in Figure 7.9a. The reconstruction accuracy in this simple example is extremely good, with only a very small bias present. The X_{max} resolution is $\sim$2.3 g/cm^2. This resolution will significantly degrade as additional shower parameters are added to the reconstruction,

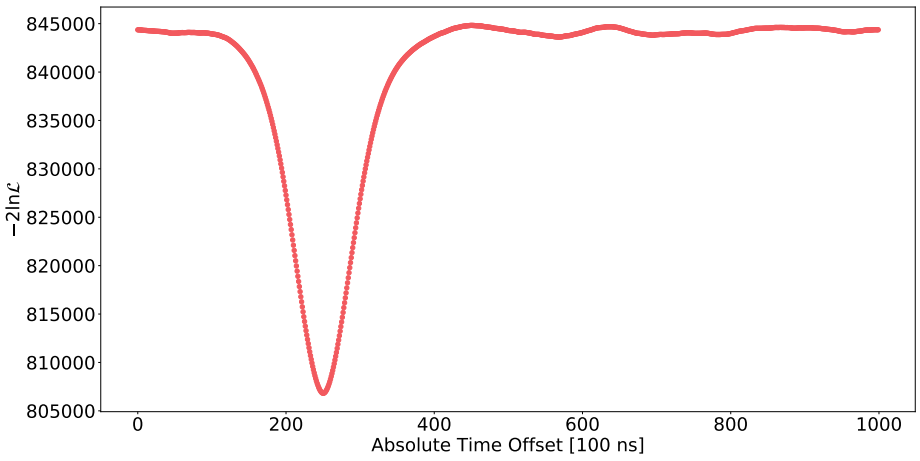


(a)

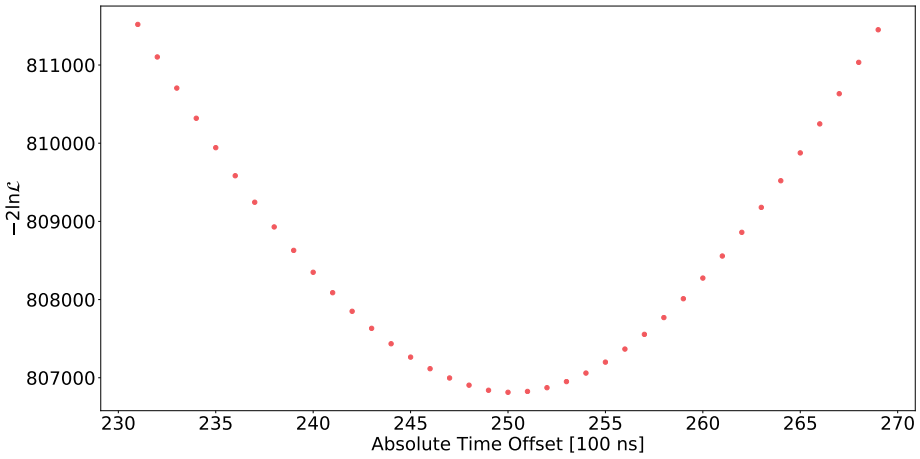


(b)

Figure 7.5. Example of PMT traces (red) from an event compared with (a) the noise-free traces from an event with all PMT traces shifted by a fixed absolute time offset of 250 time bins and (b) the comparison with the reconstructed absolute time offset applied. Note the good agreement following the reconstruction procedure. The four PMT traces are arranged based on the sky view of the simulated FAST telescopes.



(a)



(b)

Figure 7.6. The negative log-likelihood ($-2\ln\mathcal{L}$) plotted as a function of the absolute time offset for the traces shown in Figure 7.5. The full range of the trace is shown in (a), and a zoomed version is shown in (b). Note that the minimum corresponds to the 250-bin absolute time offset applied to the traces.

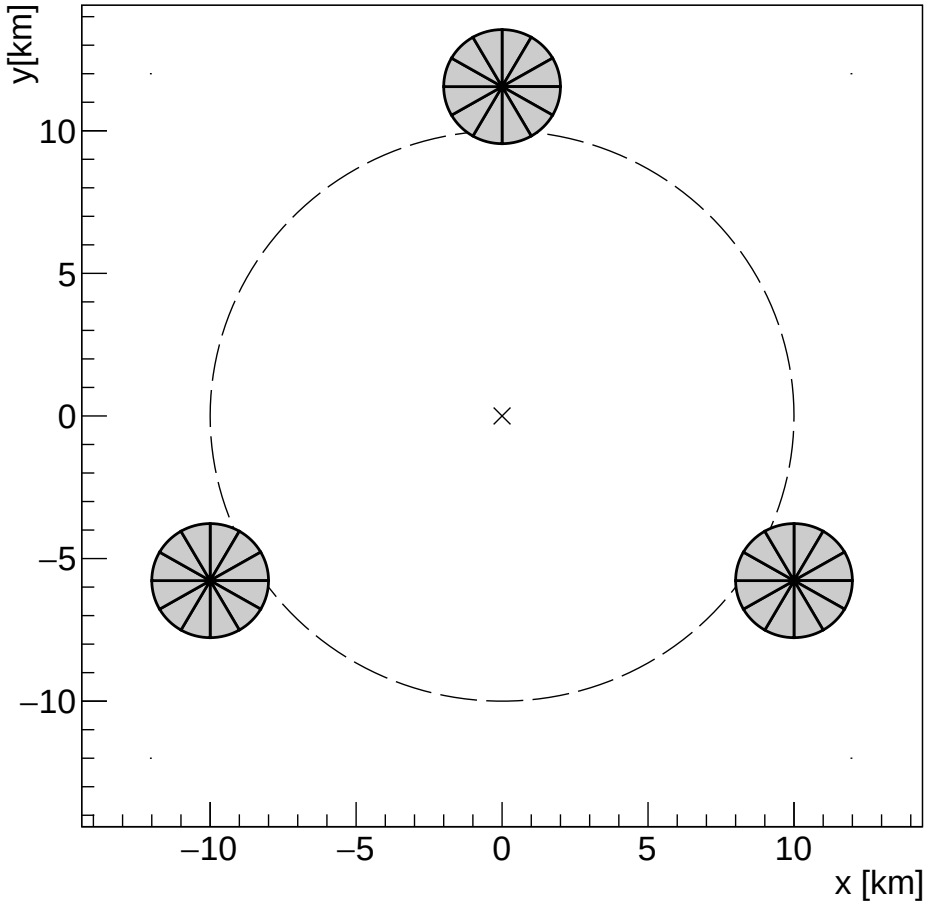


Figure 7.7. The configuration of the simulated FAST stations. The dashed line represents the region within which core positions of simulated showers are sampled uniformly, and the cross represents the centre of the cell.

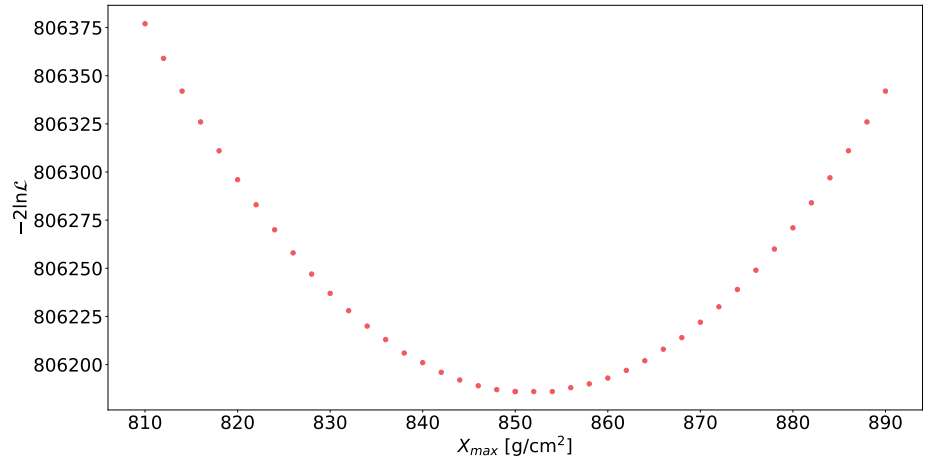
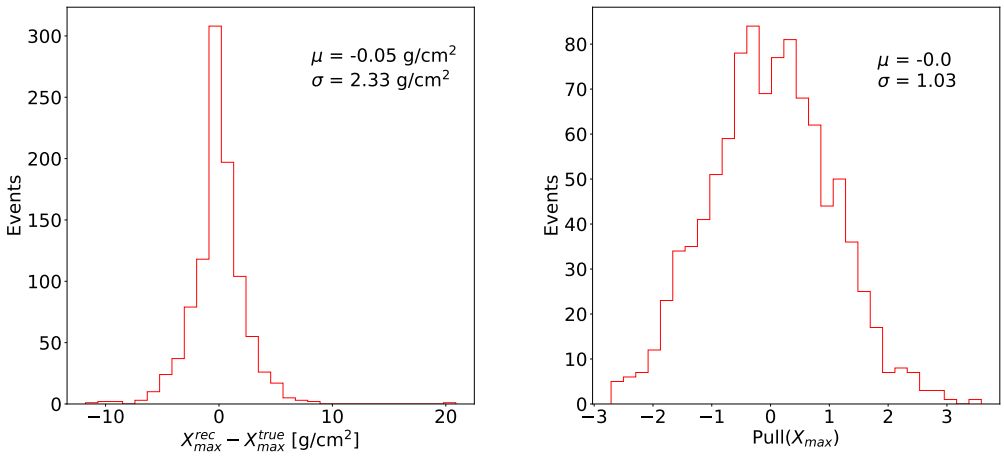


Figure 7.8. The negative log-likelihood ($-2\ln\mathcal{L}$) plotted as a function of $X_{\max}$. Note that the minimum corresponds to the reconstructed $X_{\max}$ of $\sim$852 g/cm^2.



(a) (b)

Figure 7.9. Distributions of (a) the difference between simulated and reconstructed $X_{\max}$ and (b) the $X_{\max}$ pull distribution. Note the means are close to 0 and the pull distribution is very close to a standard normal distribution (mean 0 and variance 1) indicating that the associated uncertainties in $X_{\max}$ are consistent with the random fluctuations in the reconstructed value of $X_{\max}$.

and the first guess of the shower parameters are shifted from their true values. In order to validate the uncertainties, the pull distribution, defined as

$$\text{Pull}(X) = \frac{X_{\text{rec}} - X_{\text{true}}}{\sigma(X_{\text{rec}})} \tag{7.15}$$

is shown in Figure 7.9b. The pull distribution is 3% wider than a standard normal distribution (mean 0 and variance 1). This is a negligible difference, which suggests the $X_{\max}$ uncertainties are valid for this simulated data set.

7.5.2 Energy Reconstruction

In a similar way to the $X_{\max}$ reconstruction, the value of $-2\ln\mathcal{L}$ as a function of energy is shown in Figure 7.10. The reconstructed energy for this event is 31.56 ± 0.17 EeV which agrees with the true value of 31.6 EeV within uncertainty. As for $X_{\max}$, the reconstruction is repeated for each of the 1000 simulated trial events. The distribution of the difference between the reconstructed and true energies is shown in Figure 7.11a which shows very little reconstruction bias, and an energy resolution of 0.6%. As with $X_{\max}$, this resolution is not indicative of the final energy resolution and is expected to significantly degrade as additional parameters are added to the reconstruction. Shown in Figure 7.11b is the pull distribution of the reconstructed energies. For this simulated data set, the energy uncertainties are slightly overestimated by 4% on average.

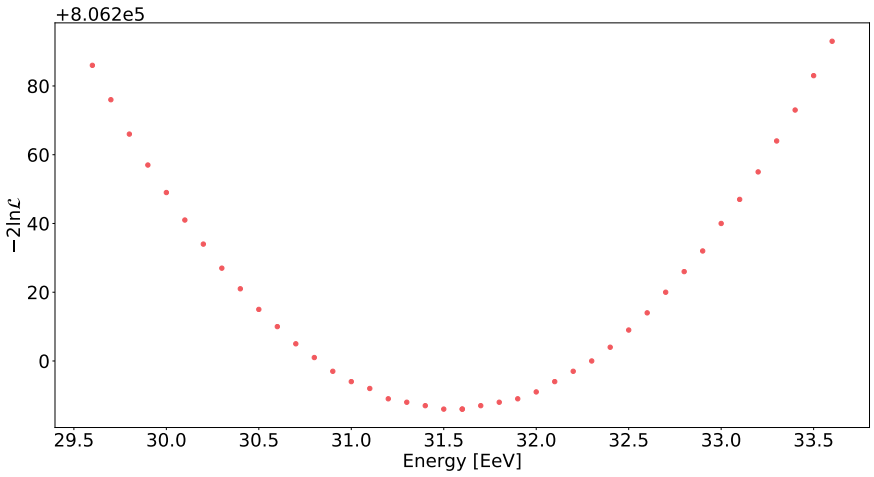


Figure 7.10. The negative log-likelihood ($-2\ln\mathcal{L}$) plotted as a function of energy.

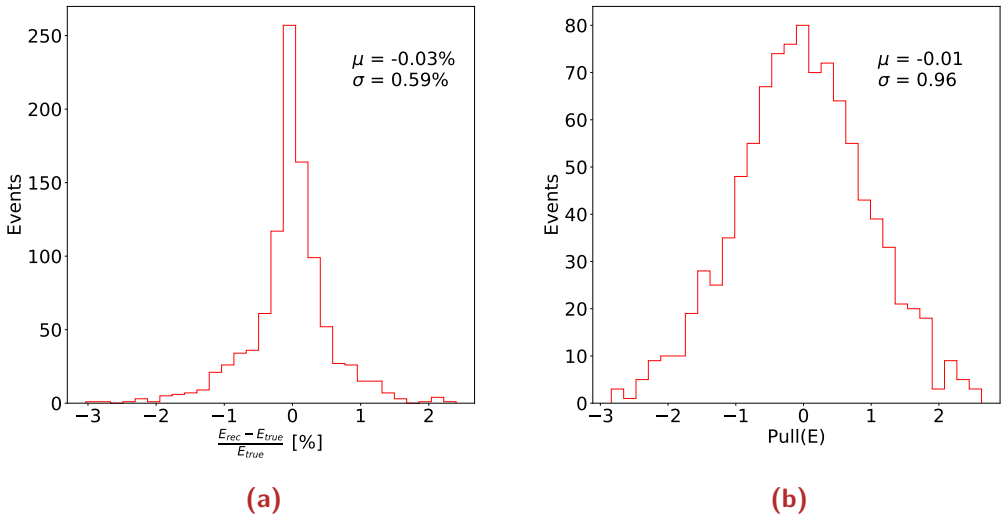


Figure 7.11. Distributions of (a) the relative difference between simulated and reconstructed energy and (b) the energy pull distribution. Note the means are close to 0 and the pull distribution is very close to a standard normal distribution (mean 0 and variance 1) indicating that the associated uncertainties in energy are consistent with the random fluctuations in the reconstructed value of energy.

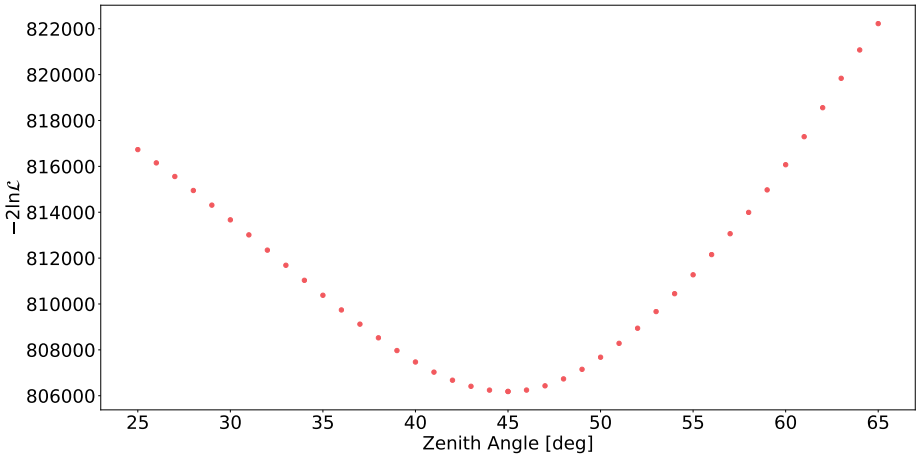
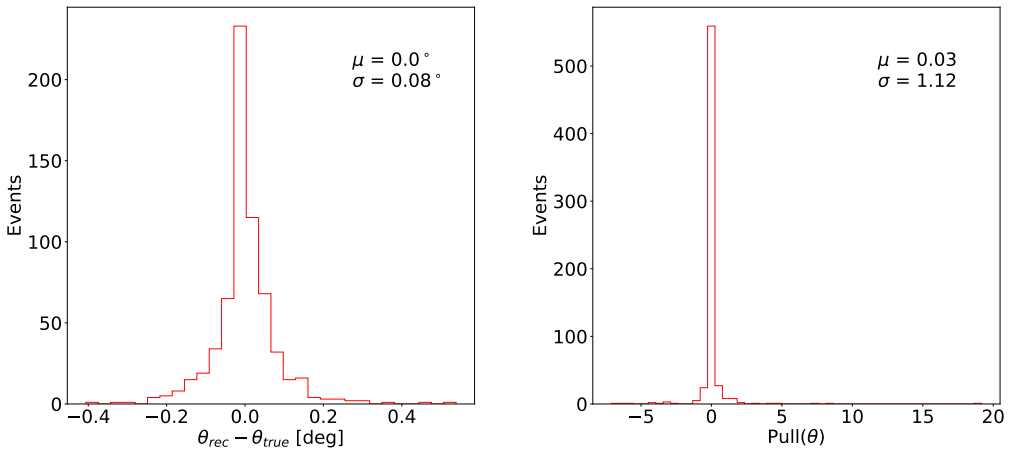


Figure 7.12. The negative log-likelihood ($-2\ln\mathcal{L}$) plotted as a function of zenith angle.

7.5.3 Geometry Reconstruction

The success of the previous tests of the energy and X_{max} reconstruction is not surprising considering that the variation of these parameters in the simulation are simple modifications to the Gaisser-Hillas profile (which is an analytical function). Other parameters are expected to be more complex due to the required transformation from signal along the shower axis to camera response as a function of time (which is not analytical). The shower geometry has a direct influence on the camera response as a function of time since a change in the position or orientation of the shower axis will change the time the signal arrives at the telescope aperture. The simplest parameter for the geometry reconstruction is the zenith angle. Unlike azimuth, changes in zenith angle represent true angles and therefore the zenith angle acts as a proxy for the reconstruction of the arrival direction.

The value of $-2\ln\mathcal{L}$ as a function of zenith angle is shown in Figure 7.12 for an example event. The reconstructed zenith angle for this event is $44.98 \pm 0.14°$ which agrees with the true value of $45°$ within uncertainty. As with the X_{max} and energy reconstruction, the zenith angle of each of the simulated trial events is reconstructed with all other shower parameters fixed to their true values. A distribution of the difference between the reconstructed and true zenith angles is shown in Figure 7.13a which shows no reconstruction bias and a resolution of $0.08°$. Shown in Figure 7.13b is the pull distribution of zenith angles. As with X_{max}, parameter uncertainties are slightly underestimated.



(a) (b)

Figure 7.13. Distributions of (a) the difference between simulated and reconstructed zenith angle and (b) the zenith angle pull distribution. Note the means are close to 0 and the pull distribution is very close to a standard normal distribution (mean 0 and variance 1) indicating that the associated uncertainties in zenith angle are consistent with the random fluctuations in the reconstructed value of zenith angle.

7.5.4 Hybrid Reconstruction

The top-down reconstruction algorithm has been able to successfully reconstruct X_{max} and energy in the simple cases considered in Sections 7.5.1 to 7.5.3. As discussed in Sections 6.1 and 7.1 FAST can operate in two distinct modes. The first being the full reconstruction using a triangular array of FAST stations which are able to measure the shower stereoscopically. The second mode relies on an independent reconstruction of the shower geometry using a SD array. In this case the FAST telescopes are only responsible for the reconstruction of X_{max} and the shower energy. This is referred to as the FAST "hybrid" mode in this work. The hybrid reconstruction mode is distinct from the conventional hybrid reconstruction, where the SD information is used along with the FD information in a simultaneous geometry fit. Instead, FAST is provided with the shower geometry as determined by an SD independently, and fits only X_{max} and energy during the top-down reconstruction.

The current configuration of the FAST prototype telescopes at TA includes three FAST telescopes which combine to provide an azimuthal FoV of $90°$. FD monocular reconstruction is particularly unreliable due to degeneracies in the orientation of the shower axis within the shower-detector plane (SDP), as well as degeneracies between the distance to the shower and its energy. Therefore, a suitable geometry reconstruction from an SD is essential for reconstruction of air showers observed by the current FAST prototype telescopes.

In order to understand the viability of the top-down reconstruction in the hybrid mode, a set of 1000 trial events have been simulated at a fixed core distance of 10 km from three FAST telescopes as shown in Figure 7.14. As for the single parameter tests, the shower energy is fixed to $10^{19.5}$ eV, and X_{max} is sampled from a parametrisation of the EPOS-LHC hadronic model. The events have been subsequently reconstructed by the top-down algorithm with only X_{max} (first guess of 850 g/cm^2) and energy (first guess of $10^{19.5}$ eV) free, and the parameters controlling the shower geometry (θ, ϕ, x, y, z) fixed to their true values. Shown in Figures 7.15a and 7.15b are distributions of the difference between the reconstructed and true X_{max} and energy, respectively. Very little bias is present in the reconstructed shower parameters. The X_{max} resolution is $\sim$23 g/cm^2 and the energy resolution is $\sim$7%. However, for these test events, the parameters controlling the shower geometry have been fixed to their true values in the reconstruction. This is impractical to achieve for measured data since the shower geometry obtained independently from an SD will only be determined within a certain resolution. The resolution of the SD must be taken into account in order to determine a realistic top-down reconstruction performance. The performance of the hybrid top-down reconstruction is discussed further in Section 9.2.1.

7.5.5 Full Reconstruction

Given the success of individual parameter reconstruction in the case of X_{max}, energy and zenith angle, this section will consider a case study of a full reconstruction where all shower parameters are free to vary. The simulated event has an energy of $10^{19.5}$ eV (31.6 EeV), an X_{max} of 750 g/cm^2, a zenith angle of 30°, and a core position close to the centre of the triangular cell shown in Figure 7.7. The Gaisser-Hillas shape parameters X_0 and λ are fixed to their average values of -121 g/cm^2 and 61 g/cm^2, respectively, since it is not expected that FAST will be sensitive enough to reconstruct them. The shower parameters of the simulated event are summarised in Table 7.2 along with the reconstructed values obtained from the top-down reconstruction with all parameters free. All shower parameters are consistent with their true value within uncertainty, validating the top-down reconstruction for this particular simulated event.

It is important to note that the true values of the simulated event were supplied to the minimisation as a first guess. This highlights an essential consideration for the top-down reconstruction. The vast phase space of the six shower parameters makes it almost impossible to rely purely on the minimisation of the likelihood function in the top-down reconstruction due to the computational requirements of performing a full detector simulation at each iteration of the minimisation procedure. An initial first guess of the shower parameters will be vital for the top-down reconstruction.

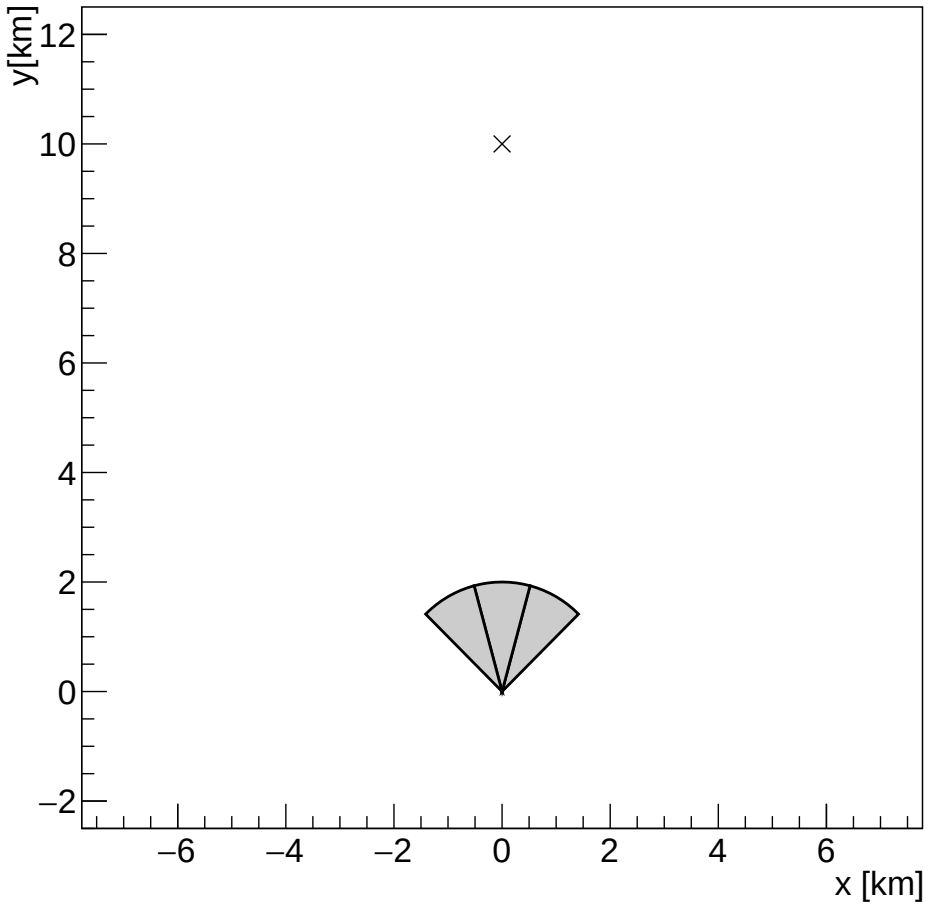


Figure 7.14. The configuration of the simulated FAST telescopes for the hybrid reconstruction tests. The cross represents the fixed core position of the simulated showers.

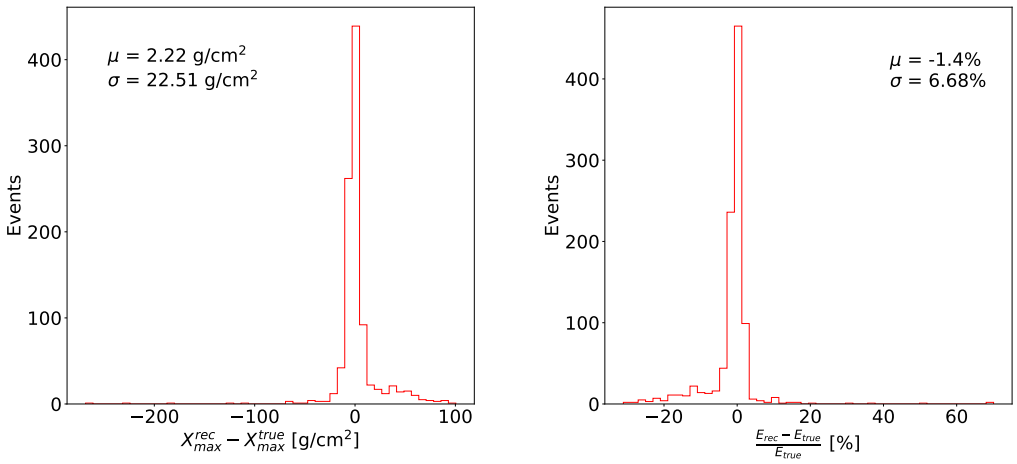


Figure 7.15. Top-down reconstruction performance for (a) $X_{\max}$ and (b) energy for simulated events at a fixed energy of $10^{19.5}$ eV and core positions as indicated in Figure 7.14.

Parameter	Simulated Value	First Guess	Reconstructed Value
$X_{\max}$ [g/cm^2]	750	750	751.8±9.6
Energy [EeV]	31.6	31.6	31.2±0.7
Zenith [deg]	30	30	31.2±0.3
Azimuth [deg]	50	50	49.8±0.8
CoreX [m]	500	500	516.0±45.6
CoreY [m]	-500	-500	-515.9±34.4

Table 7.2. Summary of the simulated and reconstructed shower parameters for an example simulated event where the true parameters are passed as the first guesses.

The more accurate the first guess of the shower parameters is, the faster, and more likely the top-down reconstruction will converge on a minimum. At this stage it is not possible to determine a realistic top-down reconstruction performance without first developing a first guess reconstruction algorithm. A first guess algorithm is considered in Chapter 8.

The event reconstruction shown in Table 7.2 has been repeated with alternate first guess parameters. Two different sets of first guess parameters are considered. In the first case (Table 7.3a), the reconstruction algorithm succeeds in recovering the true shower parameters. However, in the second case (Table 7.3b), the reconstruction algorithm is unable to recover and fails to determine accurate estimates of the shower parameters. This example illustrates the importance of a first guess for the top-down reconstruction; if the first guess of some or all of the shower parameters is too far from its true value, the top-down reconstruction algorithm fails.

Parameter	Simulated Value	First Guess	Reconstructed Value
X_{max} [g/cm^2]	750	700	754.2±9.9
Energy [EeV]	31.6	30	31.3±0.7
Zenith [deg]	30	30.5	29.2±0.1
Azimuth [deg]	50	53	50.4±1.0
CoreX [m]	500	550	455.1±7.1
CoreY [m]	-500	-350	-478.8±36.7

(a)

Parameter	Simulated Value	First Guess	Reconstructed Value
X_{max} [g/cm^2]	750	650	920.8±12.5
Energy [EeV]	31.6	28.0	24.4±0.7
Zenith [deg]	30	40	40.0±0.1
Azimuth [deg]	50	60	56.8±0.1
CoreX [m]	500	700	700.2±51.9
CoreY [m]	-500	-200	-199.8±5.3

(b)

Table 7.3. Summary of the simulated and reconstructed shower parameters for an example simulated event where the first guess of each parameter is randomly varied from the true value for (a) smaller changes from the true shower parameters where the reconstruction is able to accurately recover the true values of the shower parameters, and (b) larger changes from the true shower parameters where the reconstruction is unable to recover the true values of the shower parameters.

7.6 Conclusions

A proof-of-concept has been shown for a top-down reconstruction technique for FAST. The reconstruction algorithm has been developed and tested in this chapter. Although computationally expensive, and sensitive to the first guess of the shower parameters, the reconstruction technique has been shown to be viable for simulated events. The top-down reconstruction is a viable option for the reconstruction of air showers measured by FAST in two different detection modes: the full reconstruction utilising measurements from a triangular cell of FAST stations if paired with a suitable first guess reconstruction algorithm, and a hybrid reconstruction where only X_{max} and energy are determined by the top-down reconstruction based on the shower geometry determined by an SD. In the second case, the top-down reconstruction alone is sufficient for the reconstruction to converge on a minimum. This is due to the reduced parameter phase as well as the reduced computation time of the simulation with a smaller number of FAST telescopes. However, the resolution of the SD geometry must be taken into account in order to understand the true reconstruction performance of the hybrid mode. This is considered in Section 9.2.1.

In order to take advantage of the full reconstruction using a triangular cell of FAST stations, a reasonable first guess of the shower parameters must be passed to the top-down reconstruction in order for the minimisation algorithm to converge on the best-fit shower parameters. A first guess reconstruction algorithm has been developed in Chapter 8. Without a realistic detector simulation, the simulated PMT traces will not match those of measured events. Therefore, the success of the top-down reconstruction on measured events also depends on the quality of the simulation, including a good understanding of the detector and its response.

8 Neural Network Reconstruction

8.1 Introduction

The work presented in this chapter adopts a machine learning approach in order to reconstruct the shower geometry, $X_{\max}$ and energy of events measured by the Fluorescence detector Array of Single-pixel Telescopes (FAST). The array reconstruction mode which uses a triangular cell of FAST stations will be considered in this chapter. Typically, the shower parameters are derived using bottom-up approaches, as discussed in Chapters 3 and 7. However, with the use of machine learning techniques, the shower parameters can be extracted from several observables even though they may be weak on their own. The main motivations of this chapter follow directly from the issues which arose from the top-down reconstruction in Chapter 7 including processing time, and sensitivity to the first guess. As such, the reconstruction technique developed in this chapter will be considered as a first guess of the shower parameters for the top-down reconstruction.

A neural network has been designed to reconstruct extensive air shower (EAS) properties from measured parameters of the FAST detector. In this work, three parameters are input to the neural network per photomultiplier tube (PMT). These parameters are

- **centroid time** (signal-weighted time average) - provides information about the relative time of arrival of signals at each PMT,
- **total signal** - provides information about total signal measured from the shower as well as the relative signal between PMTs,
- **pulse height** - provides additional information about the shape of the signal pulse including asymmetry.

The following sections describe the simulation, network architecture, training, and performance of the neural network reconstruction.

8.2 Machine Learning and Neural Networks

An artificial neural network is a machine learning technique inspired by neural networks in biological systems like the human brain. A neural network consists of a series of neurons arranged in multiple layers, typically including an input, output, and multiple hidden layers. A given layer's neurons are connected to the following layer's neurons by weights. During training of the neural network, the weights are varied in order to minimise the difference between the predicted and true parameters. So, the neural network attempts to optimise the individual weights in order to accurately predict the output parameters.

The simplest case is when each neuron in a given layer is connected to all neurons in the previous layer. This is called a dense or fully connected layer. Such a system can be represented mathematically as an affine transformation connecting neurons from the previous layer to neurons in the next layer. This is represented by

$$\vec{y} = W \cdot \vec{x} + \vec{b} \tag{8.1}$$

where $\vec{x}$ represents the input from the previous neuron, W represents the weight connecting the neurons, $\vec{b}$ is a bias term, and $\vec{y}$ represents the values of the next layer of neurons. More complex neural networks employ activation functions after each layer. The weighted output of the previous layer is passed through the activation function before being passed into the next layer. This allows the network to more efficiently learn complex, non-linear structures. The system is then modified to yield

$$\vec{y} = f\left(W \cdot \vec{x} + \vec{b}\right) \tag{8.2}$$

where f represents a given activation function. There are many possible choices of activation functions depending on the application of the neural network. Sigmoid or $\tanh$ functions are popular choices for classification problems since they provide a value between 0 and 1, which can be inferred as a probability. In contrast to classification problems, where a probability is desired, the neural network in this work must predict continuous numerical values for each of the shower parameters. Therefore, in this work, where the problem is a regression one, a rectified linear unit (ReLU) activation function is used for the hidden layers, and a linear activation function is used in the output layer. The rectified linear function is given by

$$\text{ReLU}(x) = \begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases} \tag{8.3}$$

The predicted output parameters of a neural network are compared to the desired output parameters using a so-called loss function. In this work, the mean squared

error is used. Training a neural network is essentially a maximum likelihood problem. A neural network "learns" to predict the correct output parameters by maximising the likelihood using the loss function to match the output parameters with their true values. See [227] for further details of machine learning and neural networks.

8.3 Detector Simulation

It would be inefficient to consider a large array of many FAST detectors as an input to a neural network due to the complex training requirements and computational limitations. Therefore, only a representative triangular cell of FAST stations is considered in this work. This simplification reduces memory consumption and increases the neural network's ability to generalise to different sizes and configurations of a future array of FAST detectors. Even in the case where a real event is measured by a FAST array, only a triangle of stations is expected to measure the event due to the vast distances between detectors and the expected sensitivity of the FAST telescopes. In the case where a fourth or fifth station does detect some significant signal, it is likely that the event is of extremely high energy, and that the three most significant stations will provide adequate descriptions of the event for a first guess of the shower parameters. The top-down reconstruction can then include all information from any additional triggered FAST stations. It is worth noting, however, that the ideal network would be trained on the real configuration of a future FAST array to provide the best performance. The FAST telescopes are simulated using the `FAST-sim` detector simulation described in Section 7.3. The simulated FAST stations are arranged in a triangular configuration with a separation of 20 km as shown in Figure 8.1. A simple Cartesian coordinate system has been defined with its origin at the centre of the triangle. The positions the FAST stations are defined with respect to this origin. The pointing directions and fields of view (FoV) of the FAST telescopes are also shown in Figure 8.1.

As discussed in Section 7.2.2, the background noise of the FAST PMTs depends on the night-sky background (NSB) as well as the electronic noise, and the total background variance (in units of photoelectrons) is given by

$$\sigma^2_{\text{bkgd}} = n_{\text{bkgd}} \left(1 + V_g\right) + \sigma^2_{\text{elec}} \tag{8.4}$$

where n_{bkgd} is the number of photoelectrons measured from the NSB, V_g is the gain variance of the PMT and σ^2_{elec} is the electronic noise.

The total background noise, and the electronic noise were determined using data measured with the FAST prototype telescopes at the Telescope Array (TA). The electronic noise is determined by recording background data with the shutter closed. The

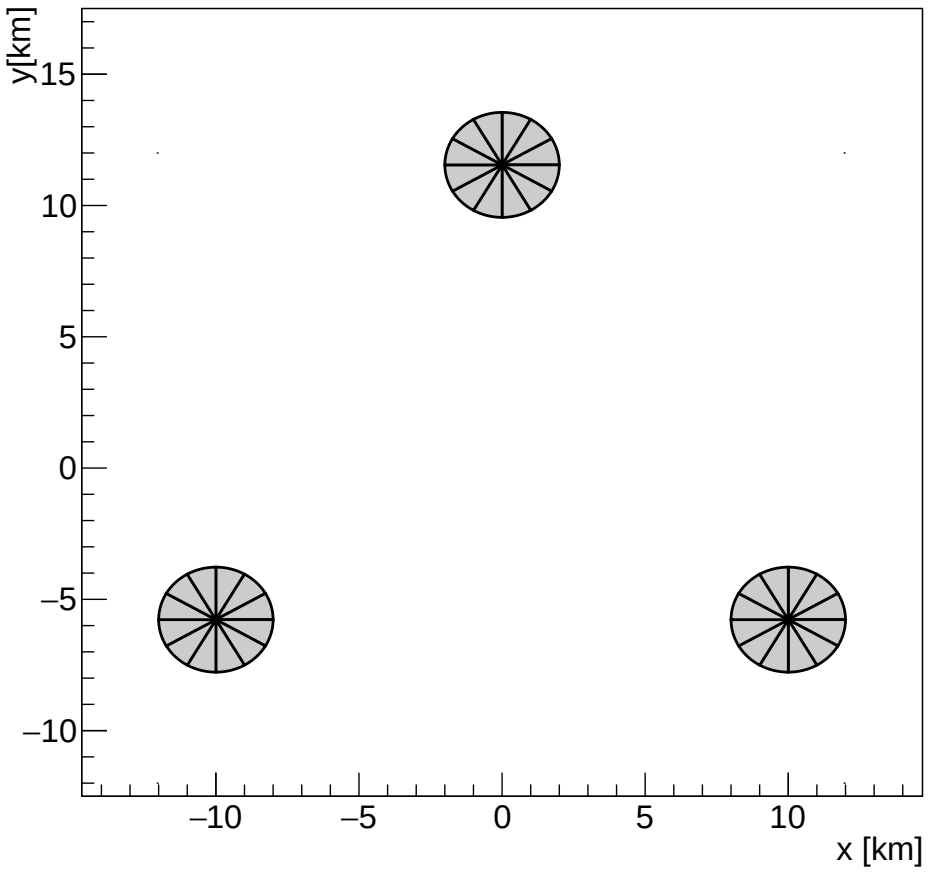


Figure 8.1. The configuration of the simulated FAST stations used for the training and testing of the neural network reconstruction.

pedestal variance is this case is ∼13 p.e./20 ns. With the shutter open, the pedestal variance increases to ∼98 p.e./20 ns [217]. This indicates that the background noise is dominated by the NSB with the shutter open. The detector simulation outputs the PMT signals in units of p.e./100 ns, so the measured pedestal variance must be converted into units of p.e./100 ns in order to determine the appropriate background variance to use for simulation of background noise. The total background variance in units of p.e./100 ns is given by

$$\sigma^2_{\text{bkgd}}\,[\text{p.e./100 ns}] = \frac{\sigma^2_{\text{bkgd}}\,[\text{p.e./20 ns}]}{\sqrt{5}} \tag{8.5}$$

since every 5 bins of the background noise are combined. This results in a background variance of ∼44 p.e./100 ns. A conservative approach is taken, where values are sampled from a Gaussian with a variance of 100 p.e./100 ns to generate the background noise of the simulated traces.

8.4 Simulated Training Data Set

When training the neural network, it is important to provide uniform sampling of the desired output parameters. Generally, this is can simply be achieved by drawing random samples of each parameter from uniform distributions. For example, if X_{max} is a desired prediction of the network, it would be most efficient to provide training samples with X_{max} drawn from a uniform distribution within some reasonable physical limits. This ensures that the network is exposed to a wide range of X_{max} values. If instead, X_{max} were sampled from a normal distribution, it is possible that the network simply "learns" to predict the mean of the distribution rather than learning useful patterns in the data in order to predict the output accurately over a large range. For this reason, sampling X_{max} and energy from a uniform distribution is a reasonable approach. Given the FAST target energy of $10^{19.5}$ eV (31.6 EeV) the energy, E is sampled from a uniform distribution between 1 and 100 EeV:

$$E \sim U(1, 100)\ \text{EeV} \tag{8.6}$$

To determine the range of X_{max} to sample, it is useful to consider the predictions of hadronic models for pure proton and iron distributions at the limits of the energy range of the simulations. The $\langle X_{\text{max}} \rangle$ of various hadronic models changes from ∼650 g/cm^2 for pure iron at 10^{18} eV (1 EeV) and ∼850 g/cm^2 for pure protons at 10^{20} eV (100 EeV). Therefore, the values of X_{max} used to train the neural network will be sampled from a uniform distribution between 500 and 1200 g/cm^2 to account for the widths of the X_{max} distributions:

$$X_{\text{max}} \sim U(500, 1200)\ \text{g/cm}^2. \tag{8.7}$$

For the shower geometry (θ, ϕ, x, y), however, uniform sampling of the individual parameters would not provide a uniform sampling of the physical phase space. The shower arrival direction (zenith and azimuth) must be sampled in such a way that they populate a hemisphere uniformly. This cannot be achieved by sampling the zenith angle uniformly, since an infinitesimal change in azimuth, $d\phi$ depends on the zenith angle. That is, the solid angle is smaller at the vertical (small zenith angle). The consequences of sampling uniformly in zenith and azimuth independently is a larger sample of arrival directions close to the vertical. Therefore, the zenith angle is sampled from a $\sin\theta\cos\theta$ distribution, which accounts for the solid angle effects, and the fact that the shower core positions are constrained to a fixed region on the ground. This also avoids extremely inclined events, which take longer to simulate due to the large atmospheric depth traversed by the shower.

The core position is sampled uniformly inside a circle, rather than from x and y independently. This ensures a uniform sampling over the entire area of the circle. To sample uniformly from a circle with centre (x_0, y_0), and radius R, it is useful to re-frame the problem in polar coordinates. An angle θ and polar distance r can then be sampled. The angle is sampled uniformly

$$\theta \sim U(0, 2\pi). \tag{8.8}$$

In order to maintain the density of points at any r, it is useful to consider the probability of sampling a point within a distance r of the centre. This is simply given by the ratio of areas if the probability is uniform. Therefore, the probability is given by

$$P(r) = \frac{r^2}{R^2}. \tag{8.9}$$

Inverse transform sampling can then be used to sample r from Equation (8.9) to obtain

$$r = R\sqrt{x} \tag{8.10}$$

where x is sampled uniformly between 0 and 1. Finally r and θ are simply converted back to Cartesian coordinates to provide the sampled core position

$$x = x_0 + r\cos\theta \tag{8.11}$$

$$y = y_0 + r\sin\theta \tag{8.12}$$

In order to sample the entire parameter phase space, the shower core positions should be randomly sampled from within the triangular cell's circumscribed circle. In the case of a 20 km spacing between the stations, this corresponds to a circle with radius $20/\sqrt{3}$ km. However, for the purpose of this work, events are restricted to three particular core areas shown in Figure 8.2; the centre of the triangle, half-

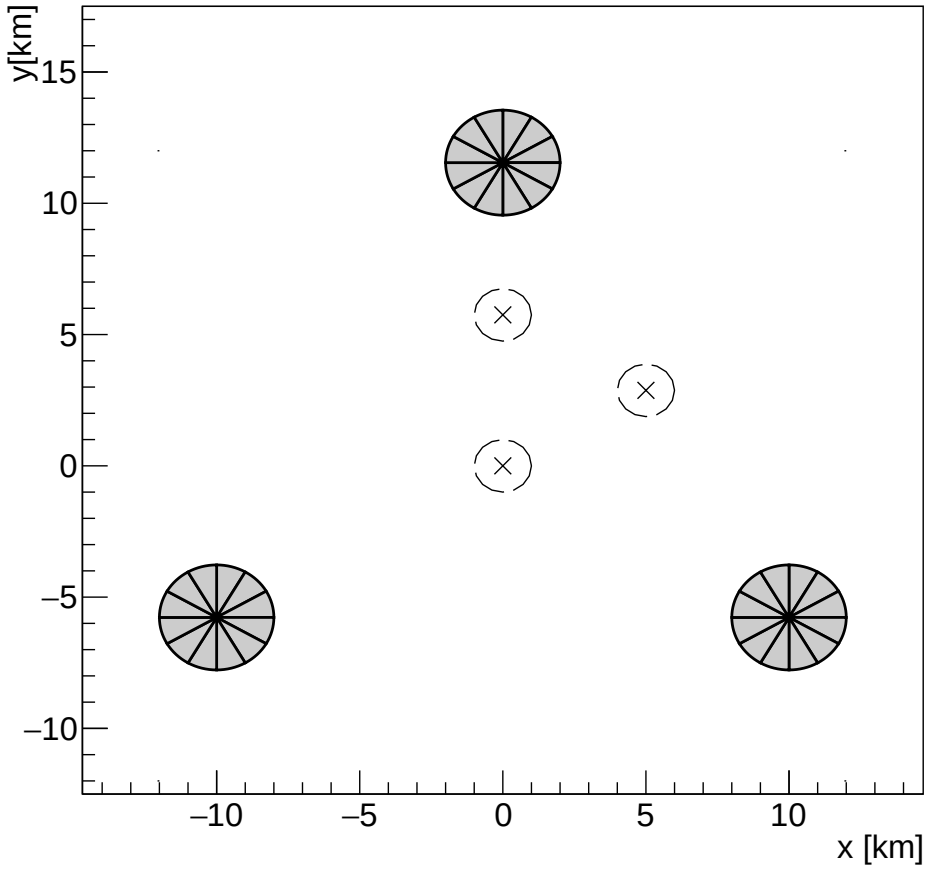


Figure 8.2. The configuration of the simulated FAST stations used for the training and testing of the neural network reconstruction. The dashed lines represent the regions within which core positions of the simulated showers are sampled uniformly, and the crosses represent their centres.

way between two FAST stations, and half-way between the centre of the triangle and the upper FAST station. These three positions, due to the symmetry of the equilateral triangle configuration of the FAST stations, represent the most extreme core locations, and can therefore be used to characterise the full parameter space.

8.5 Network Architecture and Training

8.5.1 Pre-processing

As discussed in Chapter 8, three parameters are used to characterise the pixel pulses; total signal, pulse height, and centroid time. The same procedure described in Section 3.5.2.2 is used to search for pulses in FAST pixel traces. Each trace is

searched for signals above the background noise by determining the start and stop times, k_{start} and k_{stop} which maximise the signal-to-noise ratio (SNR) in a given pixel, i:

$$\left(\frac{S}{N}\right)_i = \frac{S_i}{\sigma_i \sqrt{k_{\text{stop}} - k_{\text{start}}}}. \tag{8.13}$$

where S_i is the signal measured in the ith pixel, and σ_i is the standard deviation of the background noise estimated from early times bins without shower signal. The minimum pulse width considered is 300 ns (3 time bins). The calculation of each pulse parameter is discussed in detail below.

8.5.1.1 Signal Traces

The total signal, S, of a pulse is simply determined as the sum of the signal between the start and stop times which maximise the SNR

$$S = \sum_{j=k_{\text{start}}}^{k_{\text{stop}}} s_j \tag{8.14}$$

The signals measured by these PMTs have a large dynamic range, so it is important to consider the effect of extremely large input parameters to the neural network. Normalising the input parameters has two main benefits: it provides approximately equal weight to each input parameter and smaller changes in the input parameters improves the efficiency of the training. To normalise the total signals of each pulse the total signal is divided by the average total signal of the entire training data set. This provides a normalised total signal

$$\hat{S}_i = \frac{S_i}{S_0} \tag{8.15}$$

where S_0 is the average total signal of the entire training data set. This method preserves the relative total signals between pulses in a given event.

8.5.1.2 Pulse Height

The pulse height is defined as the maximum signal of all 100 ns bins within the pulse window. Analogously to the total signal, the pulse height is normalised by h_0, the average pulse height of the entire training data set:

$$\hat{h} = \frac{h_i}{h_0} \tag{8.16}$$

8.5.1.3 Arrival Time

Standard Fluorescence Detector (FD) reconstruction algorithms utilise the arrival times of the shower signal in each pixel to estimate the shower geometry. The FAST camera only contains four PMTs per telescope, so this information is extremely important for the reconstruction of the shower geometry. The stereoscopic measurements provide additional power to constrain the shower geometry compared with standard monocular measurements. An estimate of the signal arrival time in a given pixel can be obtained using the centroid time (signal-weighted time average) of the pulse, defined as

$$\bar{t} = \frac{\sum_{j=k_{\text{start}}}^{k_{\text{stop}}} s_j t_j}{\sum_{j=k_{\text{start}}}^{k_{\text{stop}}} s_j} \tag{8.17}$$

where s_j is the signal in units of photoelectrons in the jth trace bin, and t_j is the midpoint of the jth trace bin. In contrast to the total signal and pulse height, where the relative signals in a given event were preserved following the normalisation, the time of arrival must instead preserve the time difference between pulses in a given event. Since the absolute time of the signals are not important, the centroid times are normalised with respect to the pixel with the earliest arrival time and σ_t, the standard deviation of centroid times from the entire training data set. This results in a normalised centroid time of

$$\hat{t_i} = \frac{t_i - t_0}{\sigma_t}. \tag{8.18}$$

This normalisation is identically applied to the training data and to the independent validation and test data sets.

8.5.2 Pulse Trigger Threshold

While the neural network input parameters can be determined for any trace, whether or not it contains signal, the significance of the pulse is important for the neural network. A significance threshold can be applied in order to only retain pulses with a high significance. This has the effect of removing pulses with little or no significant signal. Pulses with a small significance are less useful since their SNRs are generally smaller, so the pulse parameters (total signal, pulse height and centroid time) are less precise, resulting in an imprecise reconstruction. In other words, input parameters which vary randomly will introduce unnecessary noise into the neural network, resulting in additional noise in the output parameters. In order to reduce such noise, only pulses above some pre-defined significance threshold should be included in the training and subsequent reconstruction. One might then decide to apply a high significance threshold cut to the pixel pulses; however, this can have the effect of removing many (useful) pulses from the event. So, a balance must be found between maintaining good quality signals, and a large enough quantity of

signals. While the purpose of this study is not to design a detector trigger algorithm, it is important to consider which pulses to include and which to disregard for the training and reconstruction. Additionally, an understanding of the limitations of the reconstruction provides important information for the future development of triggers in the detector electronics.

In addition to the precision of the reconstruction, the random trigger rate (how often triggers occur when no signal exists) is also an important measure of a reasonable significance threshold. A low threshold can be set, but many pixels will trigger randomly even though no signal is actually present. Ideally the threshold will be chosen such that only pixels with signal are selected, and all pixels with no signal are rejected. In practice this is not usually possible due to a degree of overlap in the distributions of pulses with and without actual signal.

The effect of random triggers can be determined by studying the rate of false triggers for different possible significance thresholds. A sample of 100,000 events have been simulated to study this effect. For each of the 144 pixels in an event, Equation (8.13) is used to calculate its maximum SNR. Each pixel is also labelled as either signal, or no signal based the simulated traces. A distribution of the maximum SNR of these pulses for both signal and no signal is shown in Figure 8.3. The noise distribution has a mean of 3.5σ since only the maximum SNR of each pulse is included in the distribution. Although it is entirely possible to calculate smaller, even negative values of the SNR for particular trial pulse widths, only the pulse width which maximises the SNR is chosen. The signal distribution has a peak similar to the noise peak, but an extremely long tail which represents pulses with increasing significance, well above the background noise. The two distributions clearly overlap, however, above $\sim 6\sigma$ the distribution purely consists of signal pulses. However, a significant number of pulses with signal exist below 6σ, many of which are potentially worth including in the reconstruction. The peak of the signal distribution approximately matches the peak of the noise distribution. The bulk of these pulses are signals which are well below the background fluctuations. Therefore, it is impossible to recover any useful information from them. At approximately 5σ the noise pulses have significantly fallen in number to the point where it is expected that a pulse is more likely to actually contain signal than not. Another useful test is the false positive rate, which is defined as the number of pulses for which a signal was found, but did not actually exist, divided by the total number of pulses. This is shown as a function of the SNR threshold in Figure 8.4. Based on the results from Figures 8.3 and 8.4, a threshold of 5σ is chosen for the pixel pulses used for training the neural network. Pulses with a SNR below the threshold of 5σ have all of their input parameters set to zero for the purpose of training and reconstruction. The zero parameter values are still useful, however, since they explicitly include information that the shower was not measured by particular pixels.

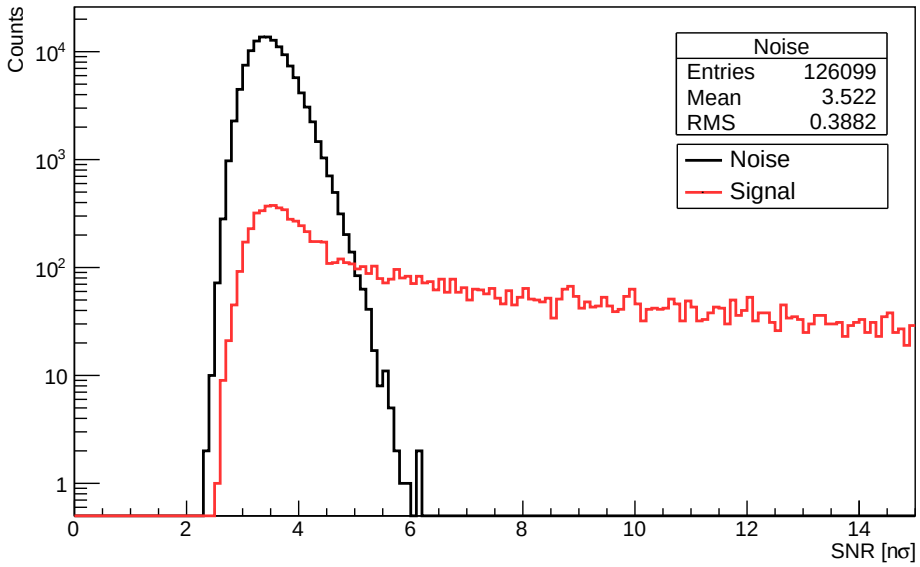


Figure 8.3. The maximum SNR for signal (red) and noise (black) pulses from simulated FAST events. Since only the maximum SNR of each pulse is included, values smaller than approximately 2σ were not obtained in this study.

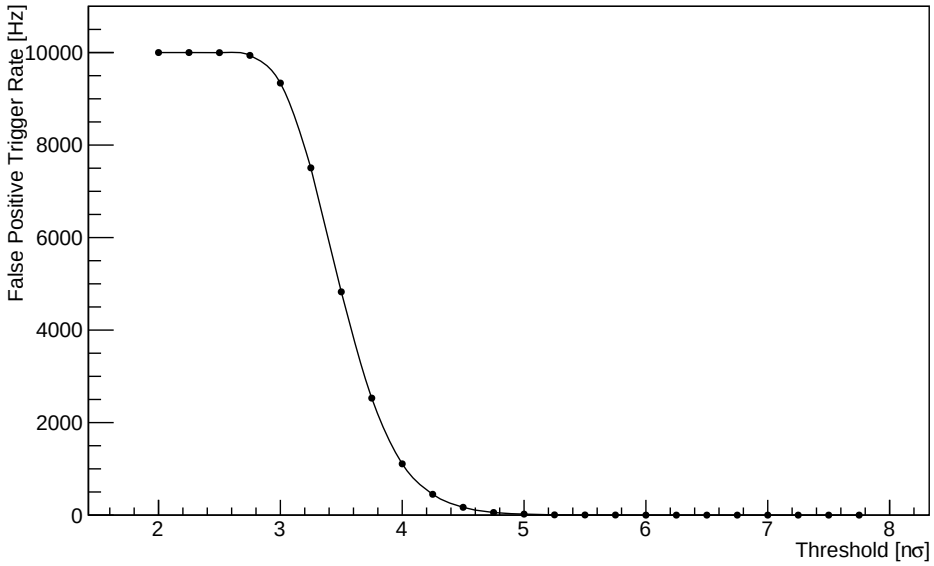


Figure 8.4. The false positive trigger rate for simulated FAST pulses as a function of the SNR threshold. This shows the fraction of noise pulses which are falsely classified as signal pulses for a given SNR threshold. Above 5σ the false positive rate is negligible.

8.5.3 Network Architecture

Three features from each pixel have been extracted from their traces in order to provide information about the signals in each PMT in the array as discussed in Section 8.5.1. Since these parameters have already been extracted from the PMT traces, a simple dense (fully connected) neural network will be employed for the reconstruction. The input to the neural network is a tensor which contains three features from each of the PMT traces for a given event; the centroid time, total signal and pulse height of each pulse. With three FAST stations in a triangular configuration, each with twelve telescopes of four pixels, this procedure provides a total of 432 individual input values to the network. These values are passed into a neural network with three hidden layers before being passed to the output layer which predicts the six shower parameters.

The network structure, including the number of hidden layers, was optimised to produce the best validation loss. Three hidden layers was chosen since adding or removing additional layers increased the validation error. Too many layers result in an extremely large number of weights which makes the neural network more computationally expensive to train. These weights must be optimised during the training procedure. A very large model can lead to over-fitting where the network "memorises" the training sample, but cannot generalise to independent data. On the other hand, a very small model will not have enough degrees of freedom to learn the complex non-linear structure of the data. Further optimisation of the neural network model should be considered in future work. One option which could be considered is a convolution neural network [227]. Such a network could potentially be used to extract features directly from the PMT traces. This would add an additional layer of complexity to the network; however, this idea may provide features which better characterise the traces. This may be beneficial to the performance of the reconstruction.

8.5.4 Training

In order to train the neural network, 500,000 events have been simulated in each of the three core regions, with the shower parameters randomly sampled as described in Section 8.4. Three separate neural networks are trained; one for each of the three core regions. The simulated data sets are separated into two samples each: 80% for the actual training, and 20% for validation. During training, a loss function based on the mean squared error of the output parameters is used. The mean squared error is given by

$$\mathcal{L} = \frac{1}{n}\sum_{i}^{n}(x_i - y_i)^2 \tag{8.19}$$

where each x_i and y_i represent the predicted and true value of each of the $n = 6$ shower parameters: X_{max}, energy, θ, ϕ, x, and y. The neural network for each of the core regions were implemented in Python using Keras [228] with TensorFlow [229]. The neural networks are trained for up to 1000 epochs. An epoch is an arbitrary cut-off defined by a single pass over the entire training data set. However, the training is stopped early if the validation loss does not decrease after 50 epochs. At the end of each epoch, the model weights are updated based on the validation loss. The ADAM optimiser [230] is used during training with a learning rate of 0.001. Further optimisation of the training parameters and settings can be considered in future work.

8.5.5 Output Parameters

During the training of the network, the output predictions of each shower parameter are compared to their true value. Since the input parameters are normalised, it is useful to normalise the output parameters of the neural network. Without normalisation, the output parameters do not have equal weighting due to the choice of loss function (Equation (8.19)). For example, the large variations in core position (thousands of meters) far outweigh the small changes in arrival direction (a few degrees). The loss function treats these parameters equally, so the training will prioritise the core position over the arrival direction, since that will lead to a smaller mean squared error. To account for this, each of the output shower parameters are normalised so that the predicted output parameters approximately lie in the range $[-1, 1]$. This is achieved using the following transformation

$$x_{\text{norm}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \tag{8.20}$$

where x represents each of the shower parameters, and x_{min} and x_{max} are the minimum and maximum values of each parameter in the entire training data set, respectively. The reverse normalisation is then applied as the final step of the reconstruction

$$x_{\text{out}} = x_{\text{min}} + (x_{\text{max}} - x_{\text{min}})\, x_{\text{norm}}. \tag{8.21}$$

In order to be consistent for the three core regions, the core position of each event is normalised using fixed values of $x_{\text{min}} = 0$ m, and $x_{\text{max}} = 12000$ m.

An exception is the output arrival direction, which is defined in spherical coordinates by two parameters: zenith angle $\theta \in [0, \pi/2]$, and azimuth $\phi \in [0, 2\pi]$. However, due to the cyclic nature of ϕ, an azimuth of $\phi = 0$ is identical to $\phi = 2\pi$. If the arrival direction is encoded purely as θ and ϕ, the fact that $\phi = 0$ is identical to $\phi = 2\pi$ is lost completely. In fact, the neural network will interpret them as significantly different values. The ideal neural network will understand that these angles are identical and

learn to predict azimuth accurately, even when close to this discontinuity in ϕ-space. In order achieve this, the geometry is converted to Cartesian coordinates

$$x = r \sin\theta \cos\phi \tag{8.22}$$

$$y = r \sin\theta \sin\phi \tag{8.23}$$

$$z = r \cos\theta \tag{8.24}$$

where $r = 1$ is the length of the arrival direction vector. The reverse transformation is applied in the final step of the reconstruction

$$r^2 = x^2 + y^2 + z^2 \tag{8.25}$$

$$\theta = \arccos\left(\frac{z}{r}\right) \tag{8.26}$$

$$\phi = \arctan\left(\frac{y}{x}\right) \tag{8.27}$$

At first glance it seems an additional output parameter has been added to the reconstruction. However, since the constraint $r = 1$ can be applied without any loss of information, the transformation can be modified to yield

$$x = \sin\theta \cos\phi \tag{8.28}$$

$$y = \sin\theta \sin\phi \tag{8.29}$$

$$z = \sqrt{1 - x^2 - y^2} \tag{8.30}$$

where only x, and y are predicted by the neural network, and z can be calculated. With this transformation, the number of output parameters is maintained, and the output arrival direction parameters lie in the range $[-1, 1]$. With this choice of encoding the arrival direction, the issue of $\phi = 0$ being identical to $\phi = 2\pi$ is circumvented.

8.6 Reconstruction Performance

In this section, the performance of the trained networks is studied using simulated events which are independent from the simulated samples used to train the networks. In order to test the reconstruction of each shower parameter, a data set of 10,000 events has been simulated. The simulated events are sampled uniformly in X_{max} and energy in order to test the reconstruction over the full training range. The reconstructed values of each parameter are compared to their true values in Figure 8.5. Gaps are present in the x and y components of the core position due to the limited range of the three core regions tested. In each case, the neural network

is able to predict the desired values as indicated by the strong correlation between the true and reconstructed shower parameters.

In order to understand any differences in the reconstruction performance of the three core positions, the energy-dependent reconstruction bias and resolution of each shower parameter have been calculated using simulated events for each of the core positions. The parameters of the simulated events are sampled from realistic distributions. The depth of shower maximum, X_{max} is sampled from parametrisations of the EPOS-LHC hadronic model. The energy of the events is simulated at fixed values from 10 EeV to 100 EeV in steps of 10 EeV. The arrival directions are sampled from a realistic $\sin\theta\cos\theta$ distribution, and the core positions are sampled uniformly within the three regions in Figure 8.2. The SNR along with the centroid time, total signal, and pulse height are calculated for each of the pixels in the three FAST stations. As for the training set, a threshold of 5σ is applied to reject noise pulses. The pulse features are passed into the neural network, and predictions of each of the shower parameters are produced. The reconstruction bias and resolution of each parameter are discussed in the following sections.

8.6.1 X_{max} Bias and Resolution

Shown in Figure 8.6 is the X_{max} reconstruction bias and resolution as a function of energy for the three core positions. The reconstruction bias shown in Figure 8.6 (left) is given by $\Delta X_{\text{max}} = X_{\text{max}}^{\text{rec}} - X_{\text{max}}^{\text{true}}$. The X_{max} bias is smallest for the central core position with a maximum bias of $\sim$5 g/cm^2, reducing to only 1-2 g/cm^2 at the highest energies. Following the central position, the upper position has the second smallest reconstruction bias which has a maximum of $\sim$10 g/cm^2 and becomes smaller at higher energy. Finally, the right core position has the largest reconstruction bias of $\sim$25 g/cm^2 at 10 EeV. There is a relatively consistent shape to the energy dependence of the reconstruction bias for each of the core positions, with a systematic shift in the X_{max} bias between the different core positions.

The X_{max} resolution shown in Figure 8.6(right) indicates a strong energy dependence with a resolution of $\sim$70 g/cm^2 at 10 EeV, becoming as small as $\sim$25 g/cm^2 at the highest energies. There is also a small core position dependence; the central position has the best resolution, followed by the right position, and finally the central position. This dependence on core position can be explained by the average distance to the three FAST stations. For the upper position, the core is a shorter distance of $\sim$5.5 km to one station, which, on average, increases the signal of one station, but decreases the signal in the other two stations which are both $\sim$15 km away. The right position follows, which is an equal distance of 10 km from two stations, and a distance of

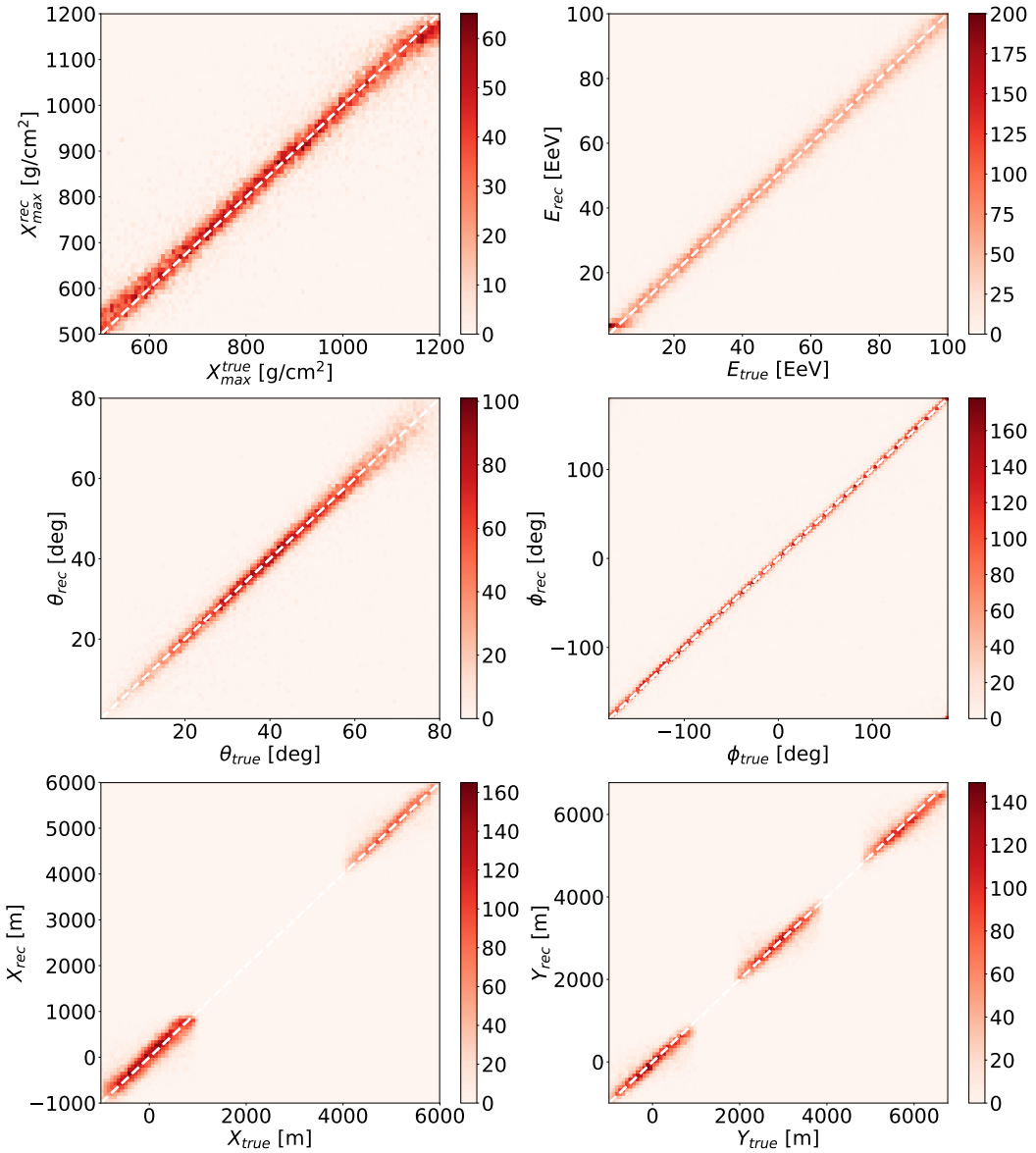


Figure 8.5. Comparison between the reconstructed and true $X_{\max}$ (top-left), energy (top-right), zenith (middle-left), azimuth (middle-right), x (bottom-left), and y (bottom-right) for all three core regions. A dashed white line is included in each panel to represent where the true and reconstructed values are equal. The gaps present in the x and y components of the core position arise due to the three chosen core regions.

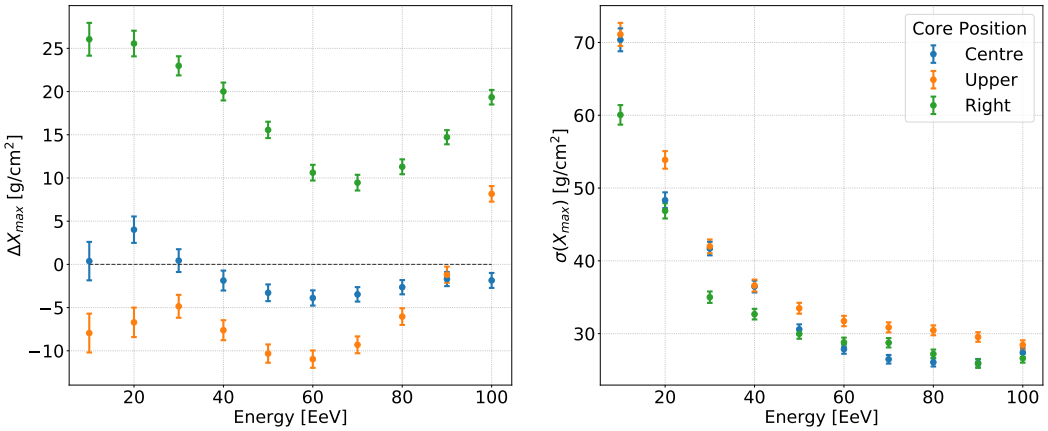


Figure 8.6. The $X_{\max}$ bias (left) and resolution (right) as a function of energy for the three tested core positions shown in Figure 8.2.

$\sim$11.5 km from the third. Finally, the central core position is an equal distance of $\sim$11.5 km from all three stations which provides the best SNR on average.

8.6.2 Energy Bias and Resolution

Shown in Figure 8.7 is the energy reconstruction bias and resolution as a function of energy for the three core positions. There is an energy dependent reconstruction bias which is reaches $\sim$17% at its largest for the right core position at 10 EeV. The reconstruction bias rapidly reduces to less than 5% at 30 EeV for all core positions. The energy resolution shown in Figure 8.7 (right) is strongly energy dependent with a resolution of $\sim$25% at 10 EeV, becoming less than 5% at the highest energies. There is little core dependence. However, the central position is slightly better, followed by the right position, and finally the upper position. This effect is likely due to the same reasons related to the signals in each station discussed for the $X_{\max}$ resolution. That is, the central core position provides the best resolution since strong signals from all three stations better constrain the energy than one much stronger signal in the upper core position.

8.6.3 Arrival Direction Bias and Resolution

As a proxy for the bias in the arrival direction, the zenith angle bias is shown in Figure 8.8 (left). The bias is very small; less than 0.5°, except for the lowest energy bin of the right core position. However, there is a statistically significant core dependence on the reconstruction bias. As for $X_{\max}$ the bias is smallest for the central core position, and larger for the upper and right core positions. The zenith angle resolution shown in Figure 8.8 (right) shows an energy dependent resolution which decreases from $\sim$4° at 10 EeV, to $\sim$2° at the highest energies. There is also

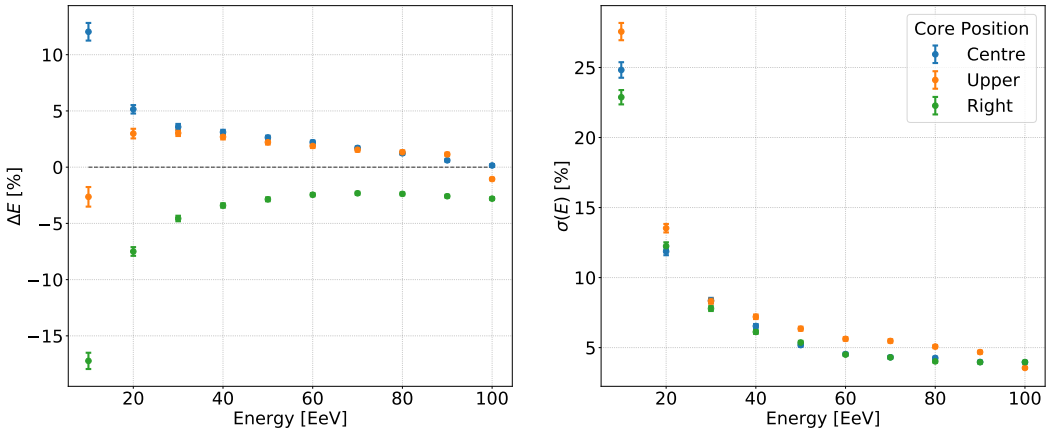


Figure 8.7. The energy bias (left) and resolution (right) as a function of energy for the three tested core positions shown in Figure 8.2.

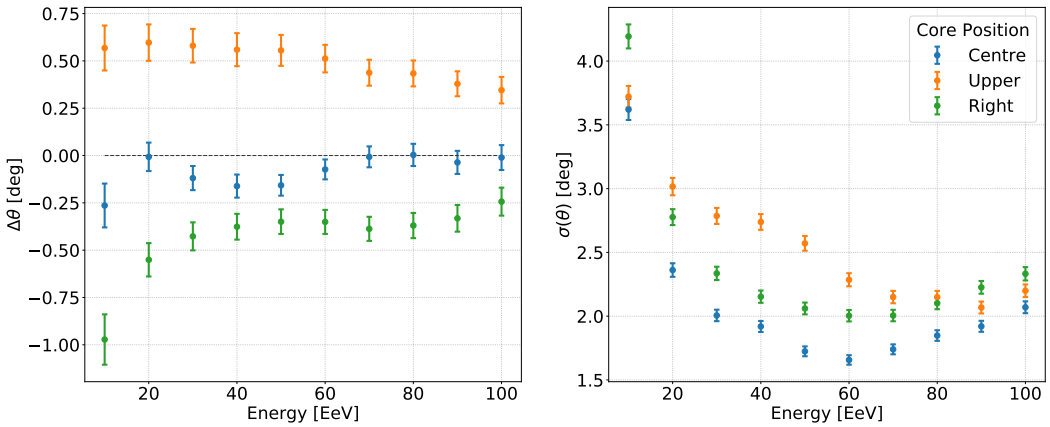


Figure 8.8. The zenith angle bias (left) and resolution (right) as a function of energy for the three tested core positions shown in Figure 8.2.

a small core position dependence in the zenith angle resolution, with the central position again performing the best and the upper position performing the worst.

The full angular resolution ψ, is defined as the angular distance containing the closest 68% of the reconstructed arrival directions with respect to their true directions. The angular distance can be characterised by the space angle, which is the angle between the vectors representing the true and reconstructed arrival directions. In spherical coordinates the space angle α, is given by

$$\cos\alpha = \sin\theta_0 \sin\theta_1 \cos(\phi_1 - \phi_0) + \cos\theta_0 \cos\theta_1 \tag{8.31}$$

where (θ_0, ϕ_0) and (θ_1, ϕ_1) are the true and reconstructed arrival direction, respectively. The space angle is calculated for each event, and a distribution of the calculated space angle is shown in Figure 8.9 for all core positions and energies.

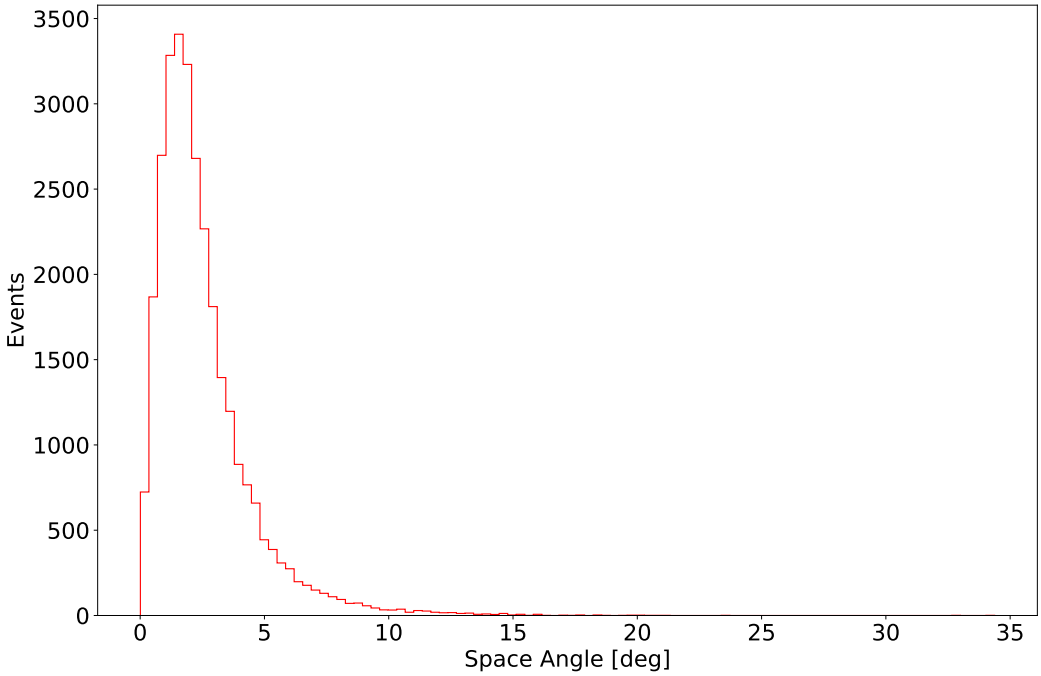


Figure 8.9. The space angle distribution for all events in the simulated test data sets.

In order to determine the angular resolution, the distribution in Figure 8.9 is integrated to determine the value of α at which the fraction of the total events equals 0.68. The cumulative distribution of space angles is shown in Figure 8.10, where the average angular resolution is determined to be 2.76°. This calculation is repeated for each individual energy bin and core position of the simulated data set, the results of which are shown in Figure 8.11. As for the zenith angle resolution, the angular resolution has a strong energy dependence. The resolution is best for the central core position, closely followed by the right core position, while the angular resolution of the upper position is slightly worse than the other two cases. Again, this can be explained by the different average core distance, and hence SNR of the pulses in the events.

8.6.4 Core Bias and Resolution

The reconstruction bias in the core position can be broken into its two x and y components shown in Figure 8.12 (left) and Figure 8.13 (left), respectively. The reconstruction bias in both components of the core position is strongly dependent on energy and core position itself. The biases in the x component of the core position are larger for the central and upper core positions compared to the bias in the right position. While the biases in the y component of the core position are smaller for the central and upper core positions, and larger for the right core position.

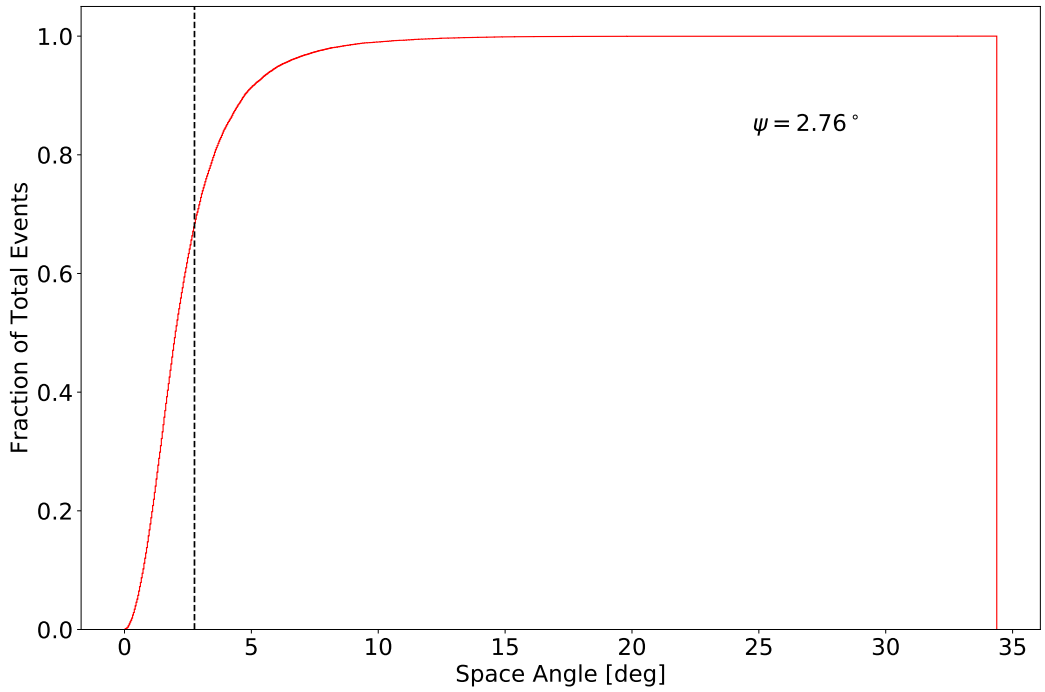


Figure 8.10. The cumulative space angle distribution for all events in the simulated test data sets. The dashed black line represents the space angle which contains 68% of all events. The average angular resolution is 2.76°.

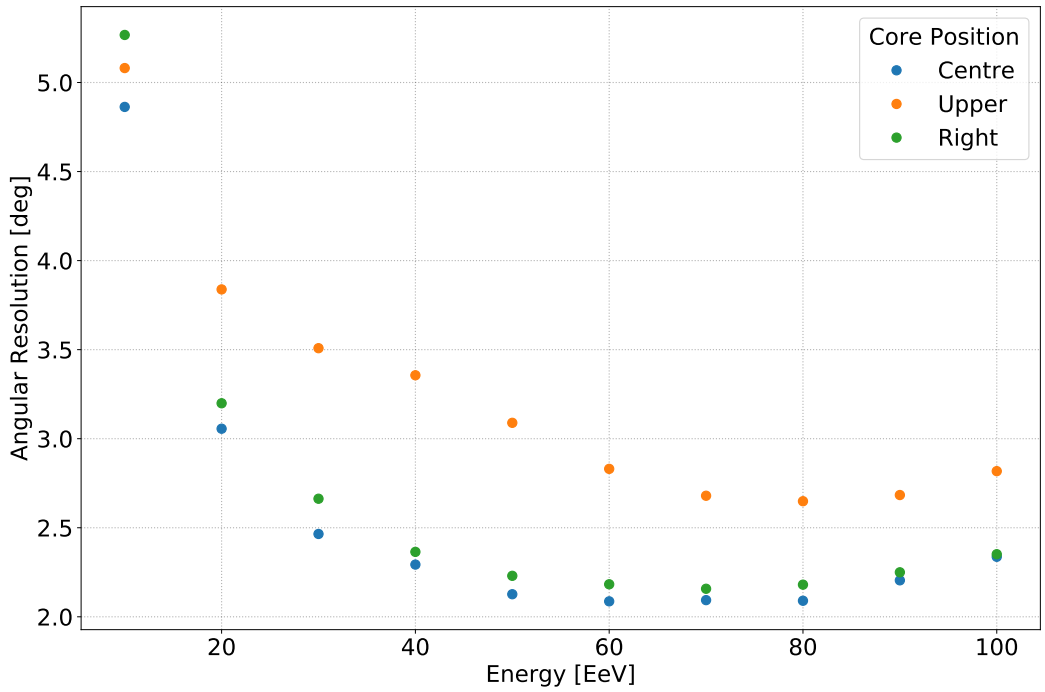


Figure 8.11. The angular resolution as a function of energy for the three tested core positions shown in Figure 8.2.

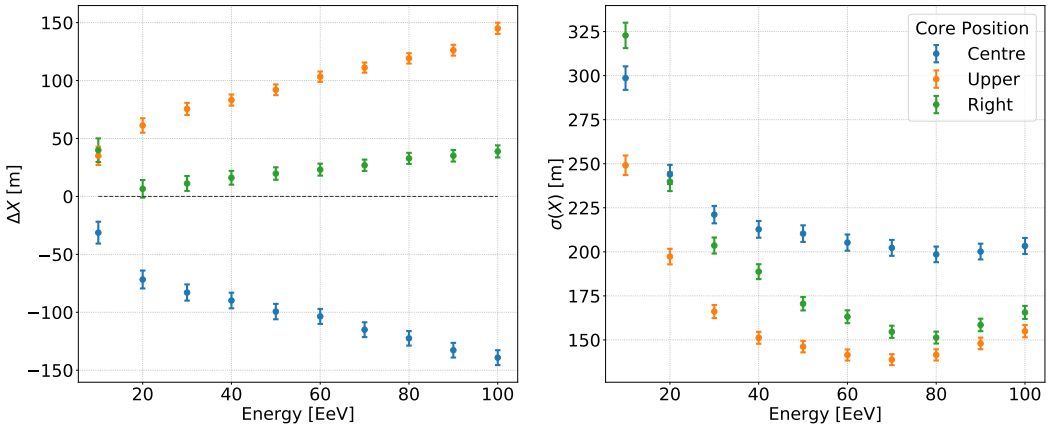


Figure 8.12. The x bias (left) and resolution (right) as a function of energy for the three tested core positions shown in Figure 8.2.

The core resolution of the individual x and y components are shown in Figure 8.12 (right) and Figure 8.13 (right), respectively. Both show a strong energy dependence, as for the other shower parameters. For the x component of the core position, the resolution is significantly better for the upper core position compared to the central core position, while for the y component of the core position, the resolution is very similar for the three core regions. Overall the y component of the core position has better resolution than the x component, with the upper core region the only exception, where the difference between the x and y components is small. This is likely due to the smaller core distances to the upper FAST station, which results in a larger SNR.

The superior performance for the y component of the core position can potentially be explained by the positions of the FAST stations with respect to the three core regions. The y component of the core position provides a good approximation to the actual core distance and is therefore strongly correlated with the SNR of each detector. As a result, the neural network is more sensitive to changes in the y component of the core position. The resolution in the x component of the core position is significantly worse for the central core region.

In a similar way to the angular resolution, the individual components of the core position can be combined based on the distance between the reconstructed and core positions d, which is simply given by the Euclidean distance between the true and reconstructed core position:

$$d = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2} \tag{8.32}$$

where (x_1, y_1) and (x_0, y_0) are the reconstructed and true core positions, respectively. The distribution of d for all simulated events is shown in Figure 8.14. To determine the core resolution, Figure 8.14 is integrated to determine the value of d at which

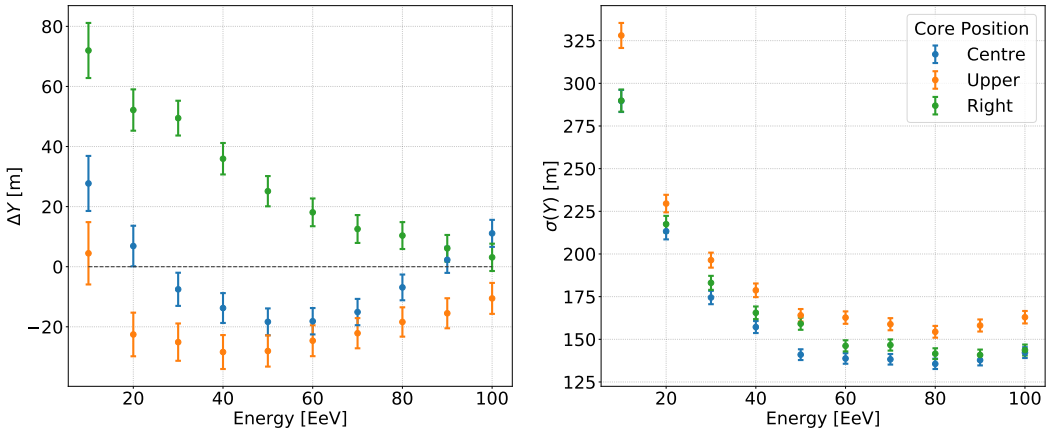


Figure 8.13. The y bias (left) and resolution (right) as a function of energy for the three tested core positions shown in Figure 8.2.

the fraction of total events equals 0.68. The cumulative distribution of d, shown in Figure 8.15, indicates an average core resolution of 267 m. The core resolution is calculated for each individual energy bin and core position of the simulated data set. The core resolution as a function of energy is shown in Figure 8.16 for the three core regions. The resolution is best for the right core region, while the difference between the central and upper core regions is negligible. A strong energy dependence in the core resolution is clear, with a resolution of ∼400 m at 10 EeV, improving to 200-250 m at the highest energies.

8.6.5 Dependence on the Number of Triggered Stations

Thus far the performance of the neural network has not distinguished between the number of triggered FAST stations or telescopes. It is expected that the performance will have some dependence on the number of triggered stations since there is a significant dependence on the core position. As discussed in Section 8.5.2, a PMT pulse is required to pass a SNR threshold of 5σ in order to be included in the neural network reconstruction. Sub-threshold pulses have each input parameter set to zero. At the highest energies, it is expected that the triangular cell of FAST stations is fully efficient. However, at lower energy this efficiency is expected to decrease.

8.6.5.1 Trigger Efficiency

To understand the energy dependence on the number of triggered stations, the trigger efficiency for only one station or a combination of only two or only three stations is calculated as a function of energy for each of the core positions. The trigger efficiency for the central core position is shown in Figure 8.17. Since the core positions, on average, are an equal distance to all three FAST stations, the difference

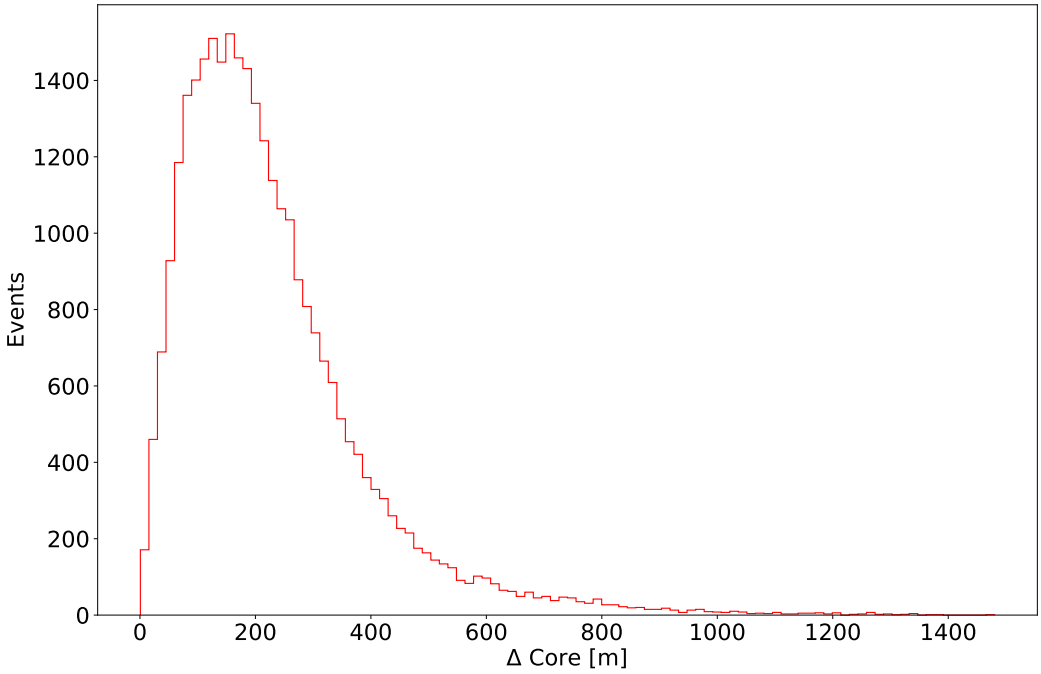


Figure 8.14. The core distance distribution for all events in the simulated test data sets.

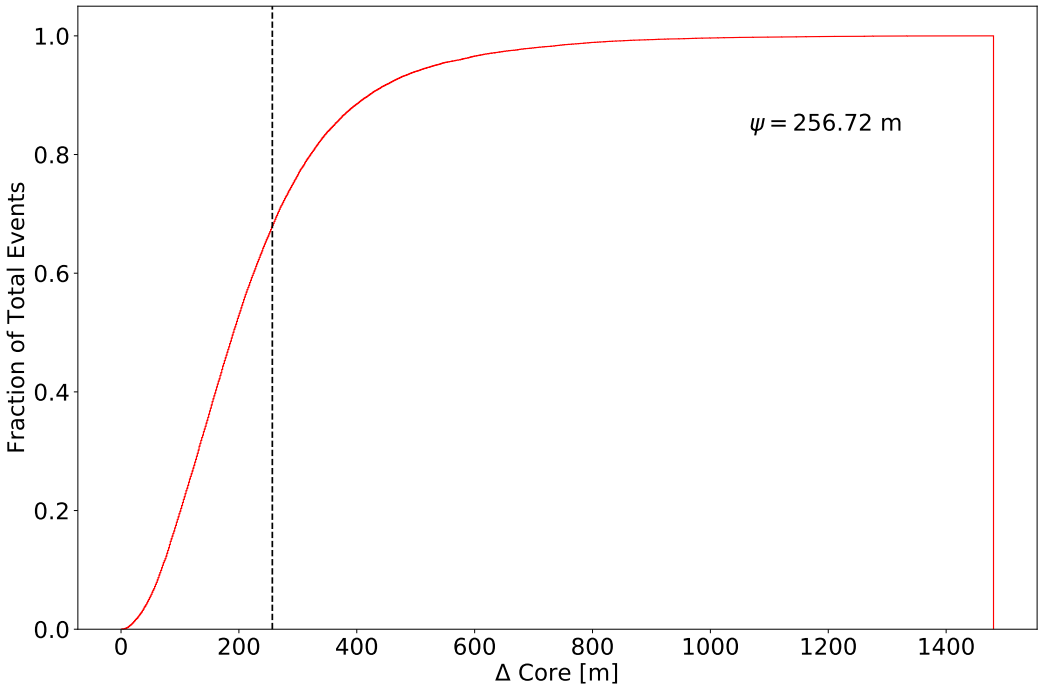


Figure 8.15. The cumulative core distance distribution for all events in the simulated test data sets. The dashed black line represents the core distance which contains 68% of all events. The average core resolution is 257 m.

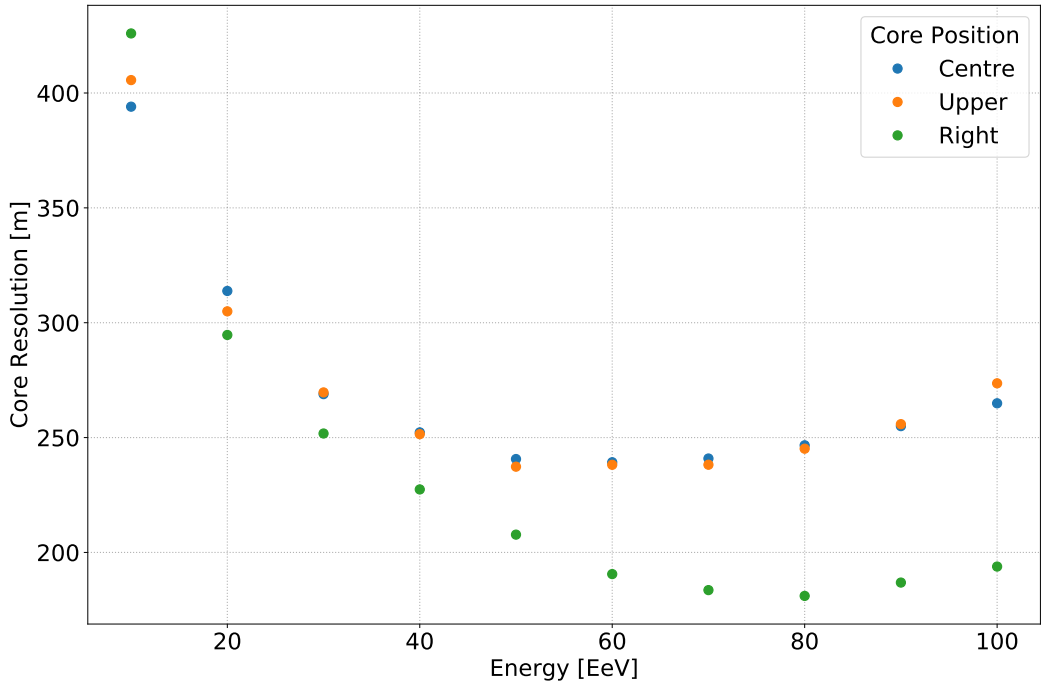


Figure 8.16. The core resolution as a function of energy for the three tested core positions shown in Figure 8.2.

between the trigger efficiency for one, two, or three stations is not significant. This can be understood by considering a vertical event at the centre of the triangular cell. The signals in each of the three stations will be identical. However, since different arrival directions are sampled from a $\sin\theta\cos\theta$ distribution for $\theta \in [0°, 80°]$, the efficiency is slightly reduced for one- or two-station triggers where some arrival directions result in the shower being physically closer to particular stations while still maintaining a central core position. The trigger efficiency for the upper core position is shown in Figure 8.18. In this case, the core positions are, on average, closer to one station, compared to the other two. This results in the trigger for a single station to stay fully efficient, even at the lowest energies. Two stations are full efficient above 10 EeV, and three stations are fully efficient above 30 EeV. The trigger efficiency for the right core position is shown in Figure 8.19. In this case, the core positions are, on average, close to two stations, and far from the third station. This results in the efficiency for a single station and two stations being approximately equal. This can be explained by the events in close proximity to two stations; once the energy of a given event exceeds ∼5 EeV, both stations are triggered with full efficiency, while the third station does not receive a large enough signal. Above ∼20 EeV the third station becomes fully efficient. For all core positions, three stations are fully efficient above ∼25 EeV. This result is consistent with the previous simulation study discussed in Section 6.3. Since the three FAST stations are fully efficient above ∼25 EeV for all core regions, the reconstruction performance is not expected to change above this energy. However, in the lowest energy bins, the trigger efficiency

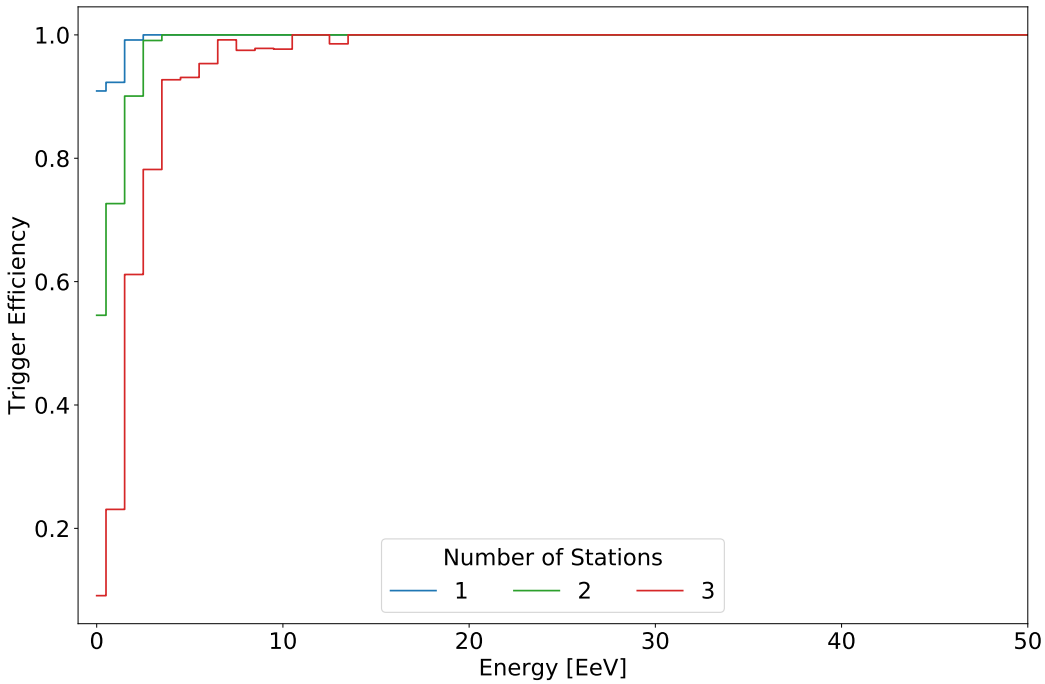


Figure 8.17. The trigger efficiency as a function of energy for the central core position for exactly one (blue), two (green), and three (red) triggered stations. The trigger is defined by a threshold SNR of 5σ as discussed in Section 8.5.2.

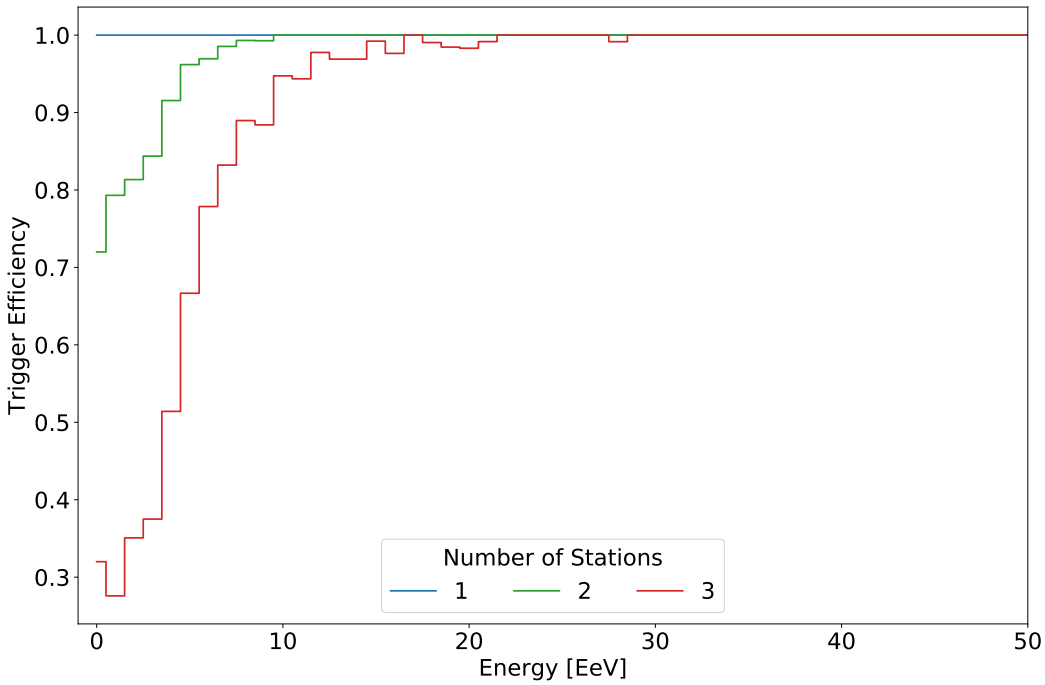


Figure 8.18. The trigger efficiency as a function of energy for the upper core position for exactly one (blue), two (green), and three (red) triggered stations. The trigger is defined by a threshold SNR of 5σ as discussed in Section 8.5.2.

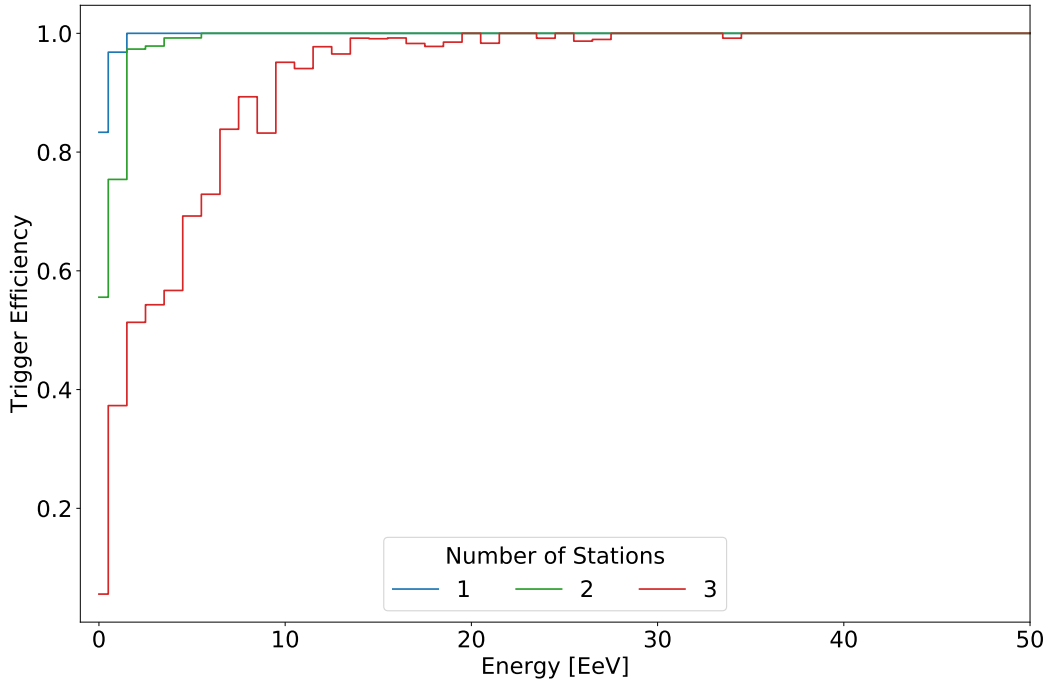


Figure 8.19. The trigger efficiency as a function of energy for the right core position for exactly one (blue), two (green), and three (red) triggered stations. The trigger is defined by a threshold SNR of 5σ as discussed in Section 8.5.2.

becomes significantly reduced, and a reduction in the reconstruction performance is expected.

8.6.5.2 Reconstruction Bias and Resolution

In order to understand the dependence on the number of triggered stations, the reconstruction resolution of each shower parameter is recalculated in the following cases: only one station, only two stations, and only three stations for 10 EeV events. For the central core region, all events at 10 EeV triggered at least two stations. Similarly, the right core region had just one event which triggered only a single station, and the upper core position had only four events which triggered only a single station. Consequently, the performance of single station events cannot be reliably determined.

The X_{max} bias and resolution is shown in Figure 8.20 for each core region as a function of the number of triggered stations. The X_{max} resolution is generally best for the three-station events followed by the two-station events. The X_{max} resolution is significantly degraded for the central core region for two-station events compared to three-station events. This is expected since the trigger efficiency for two-station events is significantly reduced at 10 EeV, as shown in Figure 8.17. However, the degradation is not as severe for the upper and right core regions. This is likely due to

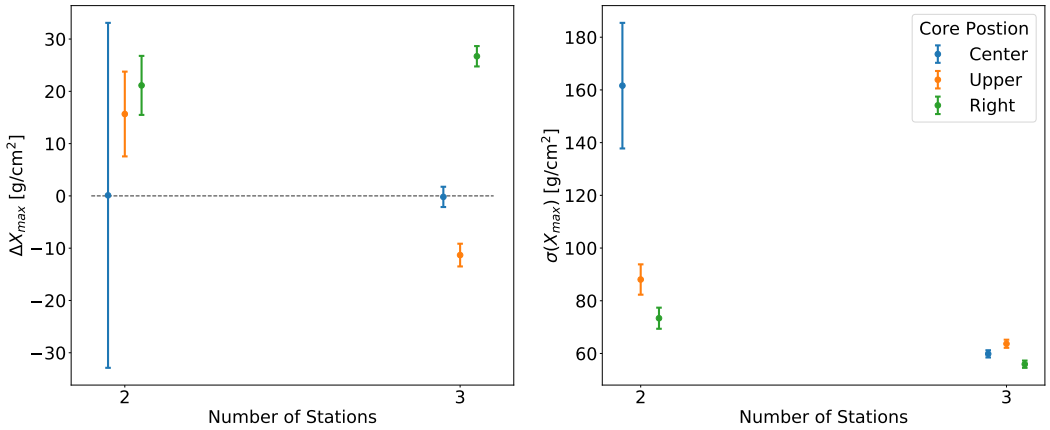


Figure 8.20. The $X_{\max}$ bias (left) and resolution (right) at 10 EeV as a function of the number of triggered eyes for the three tested core positions shown in Figure 8.2.

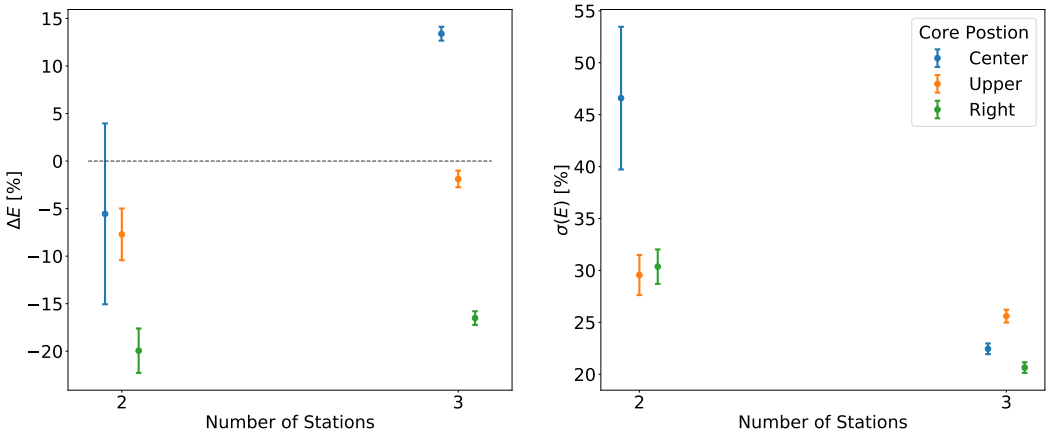


Figure 8.21. The energy bias (left) and resolution (right) at 10 EeV as a function of the number of triggered eyes for the three tested core positions shown in Figure 8.2.

the higher SNR of the pulses as a result of the smaller distance between the shower and at least one of the FAST stations.

The reconstruction bias and resolution of energy, zenith angle and core position are also shown in Figures 8.21 to 8.24, respectively. The dependence on the number of triggered stations on the other shower parameters is similar to $X_{\max}$.

8.7 Systematic Uncertainties

The performance of the neural network reconstruction thus far has only considered the ideal case where the detector state, including the atmospheric conditions, is fixed. Possible time dependent differences in the detector state include calibration changes or uncertainty, changing atmospheric conditions, and changes to the optical

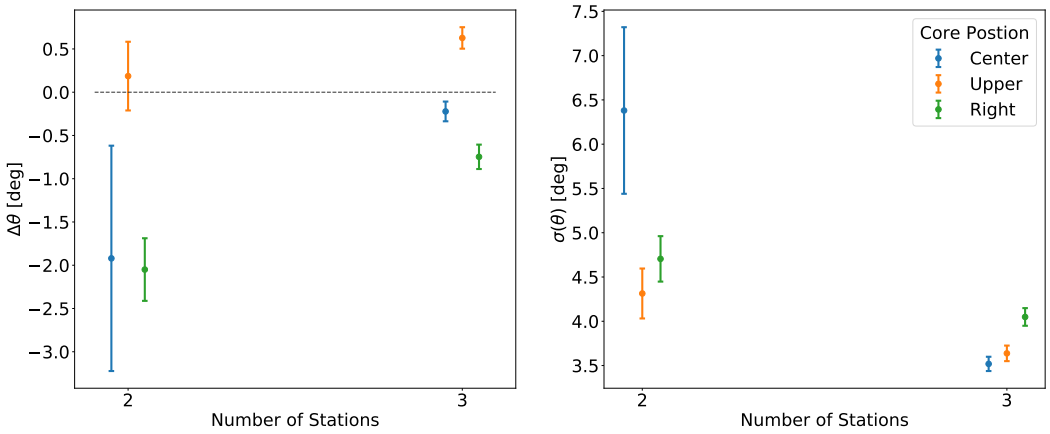


Figure 8.22. The zenith angle bias (left) and resolution (right) at 10 EeV as a function of the number of triggered eyes for the three tested core positions shown in Figure 8.2.

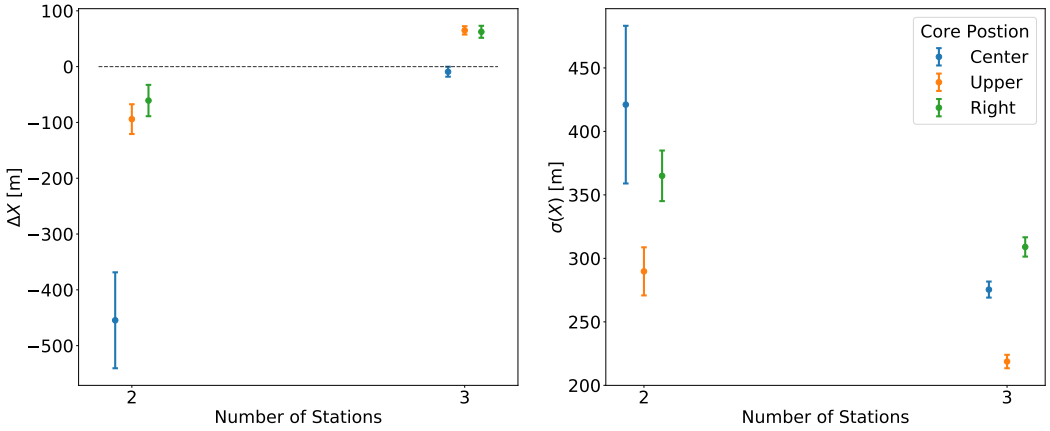


Figure 8.23. The x bias (left) and resolution (right) at 10 EeV as a function of the number of triggered eyes for the three tested core positions shown in Figure 8.2.

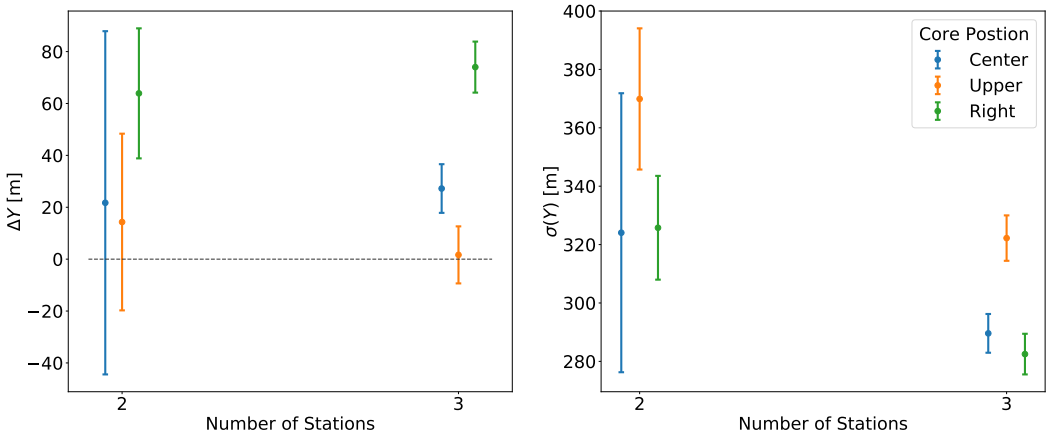


Figure 8.24. The y bias (left) and resolution (right) at 10 EeV as a function of the number of triggered eyes for the three tested core positions shown in Figure 8.2.

properties of the telescope such as mirror or filter cleaning. These effects can potentially combine to introduce additional biases in the neural network reconstruction, or to degrade the resolution. This is a potential downfall of the neural network reconstruction. In order to account for these effects, the neural network would need to be re-trained with these modified properties of the detector accounted for in the simulation. This is impractical, however, due to the large number of simulated events required to train the neural network, as well as the computational expense of training. In any case, the primary aim of the neural network reconstruction is to provide a first guess reconstruction of the shower parameters to the top-down reconstruction algorithm. If the predicted shower parameters from the neural network are systematically biased, then the top-down reconstruction will be provided with biased first guess parameters. If these biases are small, then the top-down reconstruction will be able to recover accurate estimates of the shower parameters, as discussed in Section 7.5.5. However, larger biases may cause the top-down reconstruction to fail. In contrast to the neural network reconstruction, the top-down reconstruction has the ability to account for changes in the detector state for individual events based on independent measurements of the atmospheric conditions or relative calibrations, for example. The following sections will focus on two potential systematic uncertainties in the neural network reconstruction; the telescope pointing directions, and the atmospheric conditions.

8.7.1 Telescope Pointing

One parameter associated with the detector is the direction of the optical axis (pointing direction) of each telescope. Although time-dependent changes in the telescope pointing direction are not expected, it is important to know the true pointing direction of the telescopes in order to achieve an accurate detector simulation. As part of the telescope installation procedure, the FAST telescopes are aligned, and their pointing directions are measured to a precision of $<0.5^\circ$ [231]. This is important for the detector simulation, since misalignment of the telescopes will result in very different signals at the detector, producing a bias in the output shower parameters of the neural network.

In order to investigate the effect of incorrect pointing directions on the reconstruction bias and resolution of the output shower parameters, simulated data sets have been produced with the pointing directions of the 36 telescopes in the triangular cell of stations randomly varied. For each simulated data set the pointing directions of the telescopes are shifted in a random direction in elevation and azimuth by a fixed angle. This is repeated for angles in steps of 0.2° from 0° to 1°. The simulated data sets are then processed and passed through the neural network which has been trained on simulated events with the original telescope pointing directions.

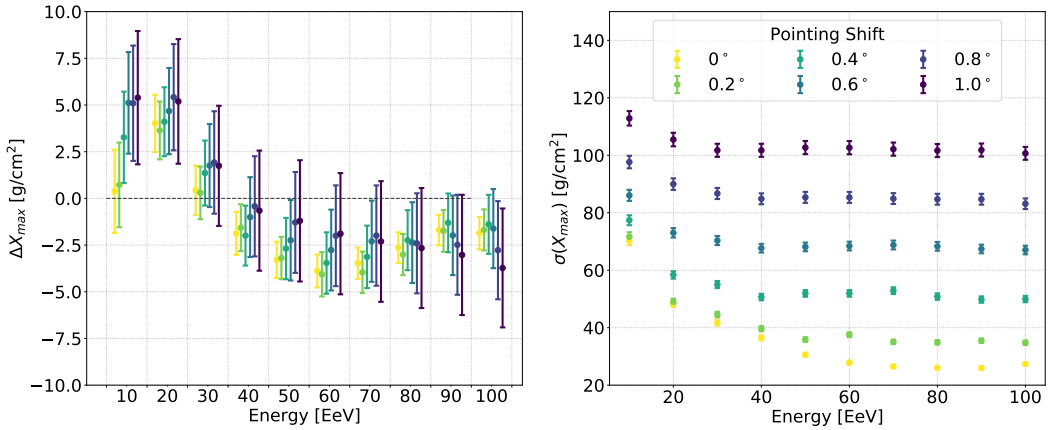


Figure 8.25. The $X_{\max}$ bias (left) and resolution (right) as a function of energy for the central core position shown in Figure 8.2 with modified telescope pointing directions. The data points in the left panel are shifted laterally to aid the reader. The grid lines separate each energy bin.

Shown in Figure 8.25 (left) is the $X_{\max}$ bias as a function of energy for the six levels of increasing pointing direction uncertainty. It is not clear if there is a significant change in the $X_{\max}$ reconstruction bias under changes in the pointing direction as the uncertainties appear to be correlated. Nevertheless, the bias is negligible. However, the $X_{\max}$ resolution shown in Figure 8.25 (right) shows a significant dependence on the pointing direction uncertainty. The randomised pointing directions of the telescopes manifest as a randomly biased estimate of $X_{\max}$ for a particular event. This has the effect of widening the distribution of the difference between the true and reconstructed $X_{\max}$, and hence degrades the effective $X_{\max}$ resolution.

The energy reconstruction bias and resolution are shown in Figure 8.26. Unlike $X_{\max}$, there is a significant increase in the reconstruction bias with increased pointing direction uncertainty. For high energy events, the reconstruction bias is small for the correct pointing directions, however, a pointing direction shift has the effect of underestimating the energy with the neural network. The energy is progressively underestimated with an increasing shift in the pointing direction. For a given shower observed by a FAST telescope, a change in the pointing direction of the telescope could have the effect of moving the shower either further into the FoV of the telescope (larger observed signal), or further outside the FoV of the telescope (smaller observed signal). For the shower energy to be underestimated on average, the signals of each shower must appear to be smaller, on average. If the pointing direction shift has the effect of moving the shower out of the FoV of telescopes, more often than into the FoV of other telescopes, then this underestimation of the energy would make sense. The energy resolution, like the $X_{\max}$ resolution, is strongly dependent of the pointing direction uncertainty. The randomised pointing directions have the effect of degrading the energy resolution, as expected.

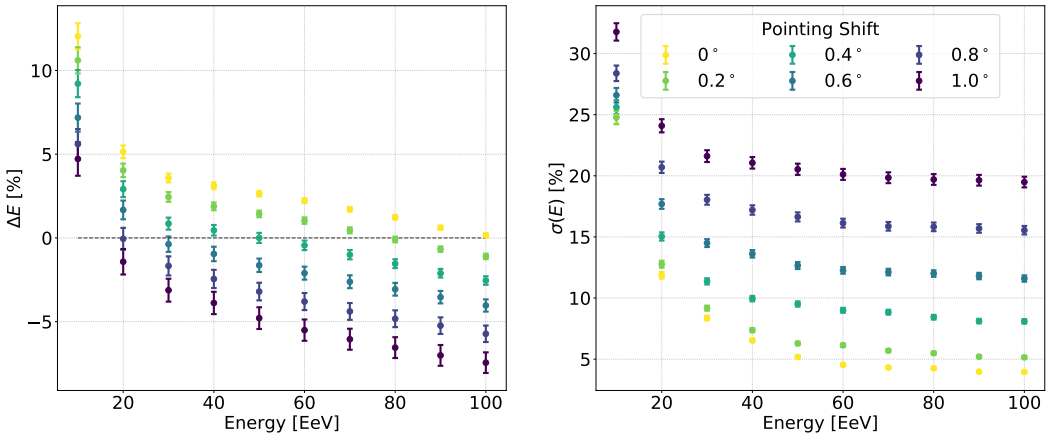


Figure 8.26. The energy bias (left) and resolution (right) as a function of energy for the central core position shown in Figure 8.2 with modified telescope pointing directions.

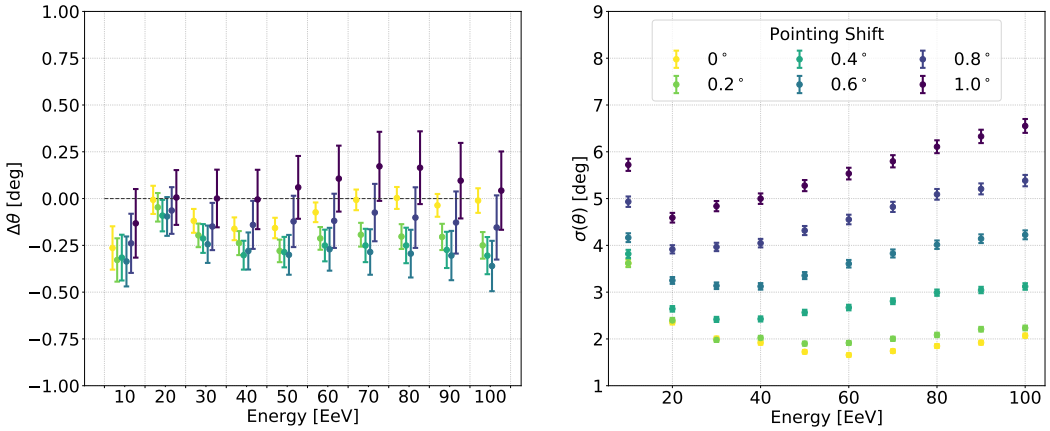


Figure 8.27. The zenith angle bias (left) and resolution (right) as a function of energy for the central core position shown in Figure 8.2 with modified telescope pointing directions. The data points in the left panel are shifted laterally to aid the reader. The grid lines separate each energy bin.

The zenith angle bias as a function of energy is shown in Figure 8.27. Any bias in the reconstructed zenith angle is negligible. However, as expected, the zenith angle resolution significantly degrades for increased shifts in the pointing direction of the telescopes. The reconstruction bias in the x and y components of the core position, shown in the left panels of Figures 8.28 and 8.29, respectively, are significantly affected by the pointing direction changes. This is likely related to the energy reconstruction bias observed in Figure 8.26, since the observed signal is dependent on both the core distance and the energy of the shower. The resolution of the core position significantly degrades with increased change in the pointing direction of the telescopes, as expected.

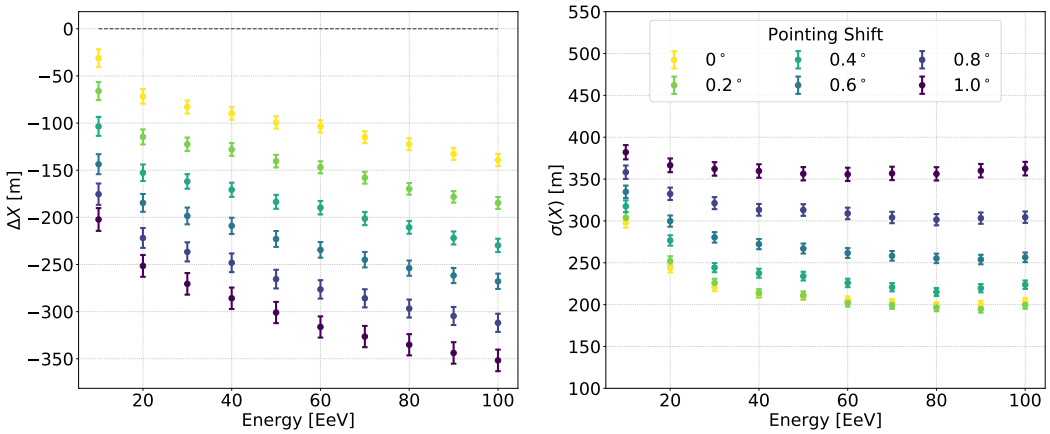


Figure 8.28. The x bias (left) and resolution (right) as a function of energy for the central core position shown in Figure 8.2 with modified telescope pointing directions.

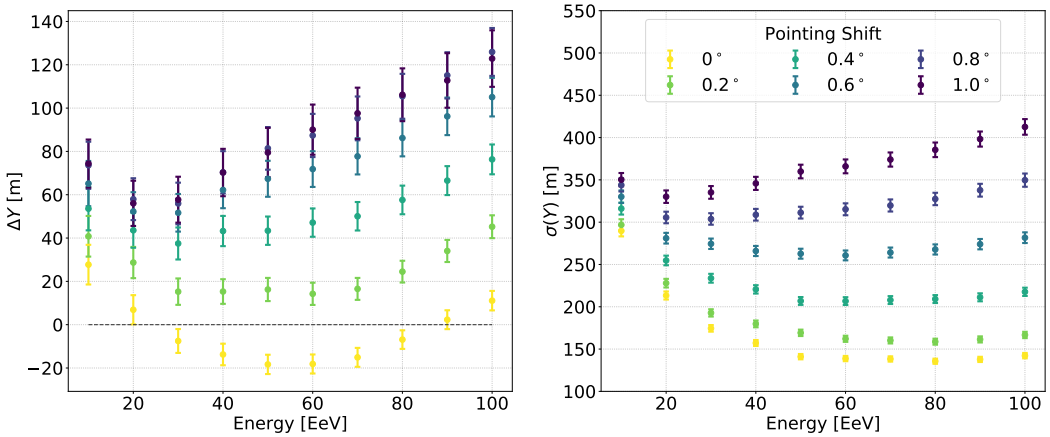


Figure 8.29. The y bias (left) and resolution (right) as a function of energy for the central core position shown in Figure 8.2 with modified telescope pointing directions.

8.7.2 Atmosphere

Another complication with the neural network reconstruction is the choice of atmospheric parameters in the detector simulation. In the FAST simulation, a parametric model with height-dependent air temperature, pressure, and humidity profiles are used to provide a realistic simulation of the atmosphere. However, both the aerosol and molecular atmospheric conditions vary from night-to-night, as well as seasonally. It is important to account for such changes in the FAST reconstruction since the atmospheric depth as a function of height above the detector strongly affects the development of the shower, and changes to the atmospheric transmission of light can significantly influence the signal received at the detector.

The top-down reconstruction can account for any atmospheric conditions by simulating changes in the atmosphere in a time-dependent way. This is provided that the atmospheric conditions at the time of a given event can be measured using atmospheric monitoring tools such as distant ultra-violet laser shots. The neural network reconstruction on the other hand, is trained on a fixed atmosphere, so changes in atmospheric conditions cannot be accounted for without re-training the entire network on a new simulated data set. It may be possible to apply corrections to the data which would account for different atmospheric conditions prior to feeding it through the neural network, however this is beyond the scope of this work, but could be considered in future work. Given that the neural network's main purpose is to provide a first guess of the shower parameters to be fed into the top-down reconstruction, it is not necessarily important to account for these systematic effects. Therefore, the focus here is to understand the sensitivity of the neural network reconstruction to the atmospheric conditions rather than find a robust solution to the problem.

Vertical profiles of the molecular atmosphere based on time-averaged monthly models have been used by the Pierre Auger Observatory (Auger) in the past [157, 232]. These monthly models include average values for the atmospheric depth, density, pressure, temperature and humidity as a function of altitude. Importantly, changes to the molecular atmosphere have the effect of increasing or decreasing atmospheric depth at a given height. This is expected to have a significant systematic effect on X_{max}. In order to understand the effects of the molecular atmosphere on the reconstruction, the monthly models are used to simulate alternate molecular atmospheres. Figure 8.30 shows the difference between the atmospheric depth of the monthly models and the molecular atmosphere used to train the neural network. The January and August models are the most extreme, therefore they are chosen as representative changes to the molecular atmosphere.

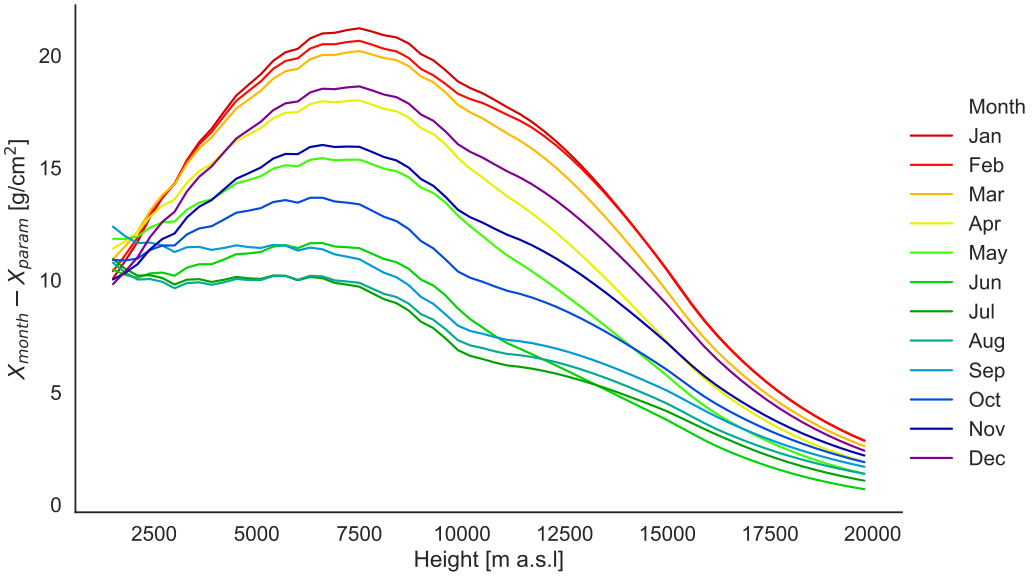


Figure 8.30. The difference in atmospheric depth as a function of height above sea level between the monthly models and the parametric model used for training the neural network. Note that January and August are the most extreme months.

In contrast to the molecular atmosphere, the aerosol atmosphere follows a simple exponential model, with the volume scattering coefficient α being described in terms of a horizontal ground-level attenuation length L, and a scale height H

$$\alpha(h) = \frac{1}{L} \exp\left(-\frac{h}{H}\right). \tag{8.33}$$

The values of H and L can be changed in the FAST simulation to provide a modified aerosol atmosphere. The vertical aerosol optical depth (VAOD) is given by

$$\text{VAOD}(h) = \int_0^h \alpha(h')dh' = -\frac{H}{L}\left[\exp\left(-\frac{h}{H}\right) - 1\right] \tag{8.34}$$

so the VAOD at infinity can be expressed as

$$\text{VAOD}(\infty) = \frac{H}{L} \tag{8.35}$$

So, it is possible to change the total VAOD in the simulated atmosphere by changing one or both of the horizontal attenuation length or the scale height. For the purpose of this study, two such atmospheres are chosen; one with a smaller aerosol content, and one with a larger aerosol content. The scale height is fixed at 3 km in both cases, and the horizontal attenuations lengths are chosen to be 300 km and 30 km, providing a total VAOD of 0.01 and 0.1, respectively. In comparison, the fixed atmosphere used to train the neural network used a scale height of 3 km and horizontal attenuation length of 53 km, providing a total VAOD of $\sim$0.055.

It is expected that an event measured under different atmospheric conditions to that with which the neural network was trained on, will be reconstructed with a bias in some or all of the output shower parameters. In order to determine the size of this potential bias, the four modified data sets described above were simulated and processed through the neural network. For the purpose of this study, the core positions were restricted to the central position (within 1 km). The reconstruction bias and resolution of each shower parameter are calculated as a function of energy for each modified atmosphere.

8.7.2.1 $X_{\mathbf{max}}$ Bias and Resolution

Figure 8.31 shows the X_{max} reconstruction bias and resolution. The resolution is relatively unaffected by changes in the atmosphere, however the data set with the increased VAOD produces the largest degradation in resolution. The reconstruction bias with the parametric atmosphere (used to train the neural network) shows some energy dependent structure as discussed in Section 8.6. This structure is maintained in the reconstruction bias for the alternate atmospheric models.

The X_{max} bias is most sensitive to the changes in the molecular atmosphere. This can be explained by the significant change in atmospheric depth of the monthly models shown in Figure 8.30. Such changes shift the position of X_{max} in height relative to the detector. Since the neural network is based on the total signal, pulse height and centroid time of each pulse, changes to the height of X_{max} relative to the detector will manifest as changes in the relative signals between pulses. The bias when using the August model is less than that of the January model since the August model is more similar to the parametric model as shown in Figure 8.30. The bias has the effect of underestimating the value of X_{max}, that is, X_{max} appears higher in the atmosphere relative to the detector. This is because the neural network is trained on an atmosphere which, at a given height, has a shallower atmospheric depth than that of the simulated test events. This is illustrated in Figure 8.32.

8.7.2.2 Energy Bias and Resolution

The energy reconstruction bias and resolution are shown in Figure 8.33. In contrast to X_{max} bias, the energy reconstruction bias is relatively unaffected by the molecular atmospheric model. However, the energy bias is strongly dependent on the aerosol atmosphere. The neural network has been trained on simulated events with a VAOD of 0.05. Reconstructing simulated events with a smaller VAOD than which the neural network was trained on, causes an overestimate of the energy. This is expected because a smaller VAOD corresponds to less atmospheric attenuation between the shower and the detector, resulting in higher signals at the detector. The neural

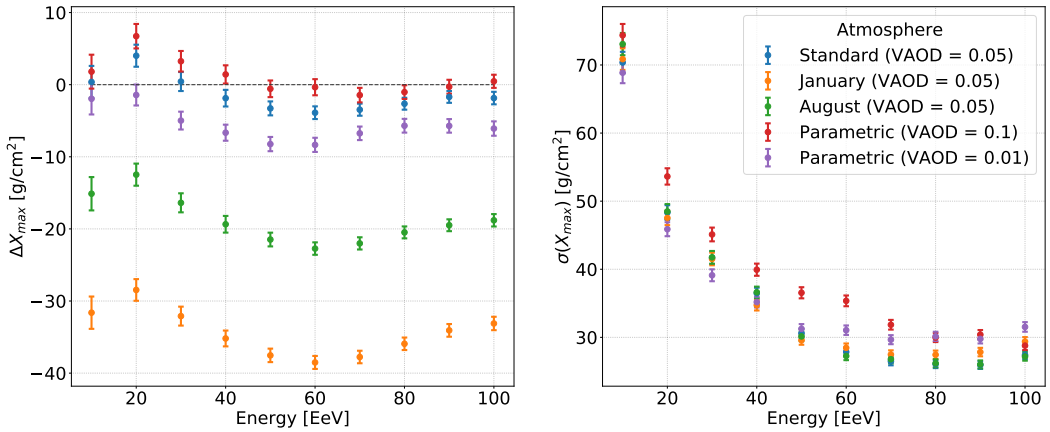


Figure 8.31. The $X_{\max}$ bias (left) and resolution (right) as a function of energy for the central core position shown in Figure 8.2 with modified atmospheric models.

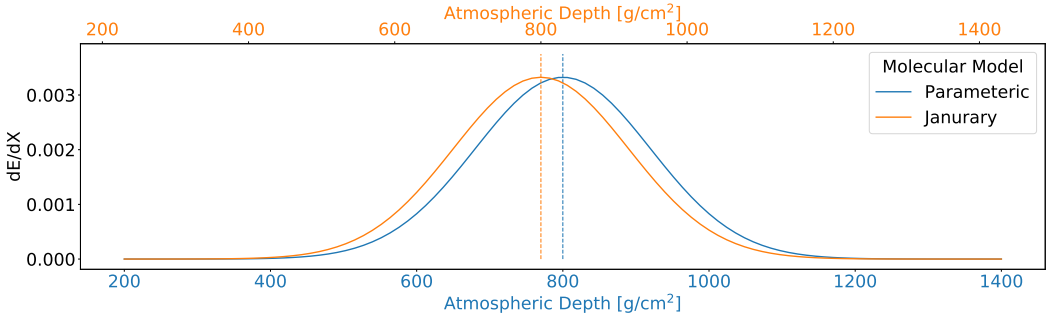


Figure 8.32. A diagram of a energy deposit profile in two different molecular atmosphere models. Shown in blue is the original parametric model and shown in orange is the January monthly model. The January molecular model is deeper in atmospheric depth at a given height above the detector compared to the parametric model (as shown in Figure 8.30). Therefore, a shower with $X_{\max} = 800$ g/cm^2 in the January model will develop higher in the atmosphere relative to the detector. This manifests as an underestimation of $X_{\max}$, as the neural network has been trained on the parametric model. So, the shower will appear at a shallower atmospheric depth corresponding to the parametric model it has been trained with.

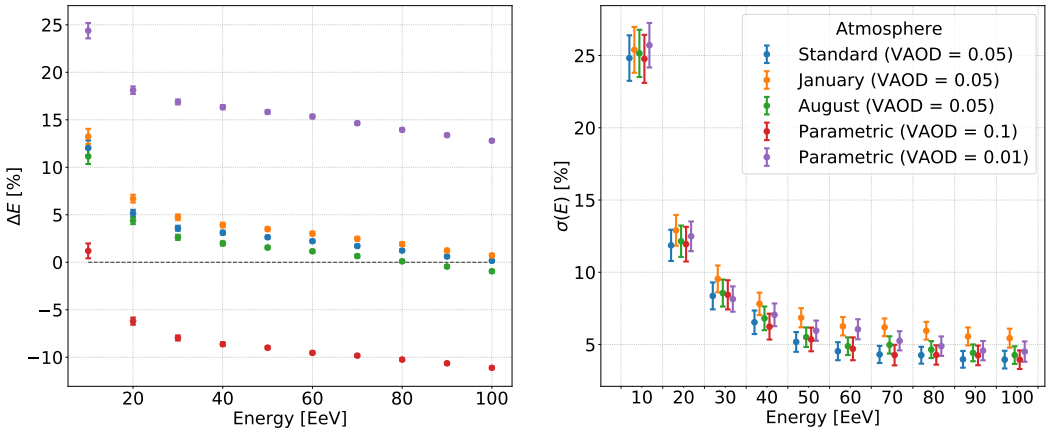


Figure 8.33. The energy bias (left) and resolution (right) as a function of energy for the central core position shown in Figure 8.2 with modified atmospheric models. The data points in the right panel are shifted laterally to aid the reader. The grid lines separate each energy bin.

network interprets these higher signals as a higher energy event. While in the case of a larger VAOD, the energy is underestimated for the opposite reason. The energy resolution, like the X_{max} resolution, is relatively unaffected by the modified atmospheric models.

8.7.2.3 Zenith Bias and Resolution

The zenith angle bias or resolution shown in Figure 8.34 do not show significant dependence on the atmospheric model. This is expected since the neural network has likely learned to use the relative timing of the pulses to predict the arrival direction of showers. The relative timing of the signals will not be significantly affected by changes to either the molecular model or the aerosol model. It is possible that the pulse shapes are altered by the modified atmospheric models, which in turn, may slightly modify the centroid time of the pulses. This can explain the minor differences observed between the different atmospheric models in Figure 8.34.

8.7.2.4 Core Bias and Resolution

The reconstruction bias and resolution of the x and y components of the core position are shown in Figures 8.35 and 8.36, respectively. Like the zenith angle, there is no significant dependence on the atmospheric model, especially for the x component. There are small differences in the reconstruction bias for the y component at the highest energies, such that the atmosphere with the smaller VAOD is more biased. This dependence on the core position can likely be explained by a weak correlation between energy and core position due to the correlation of signal and core distance. This correlation would be more likely to affect the y component for the central

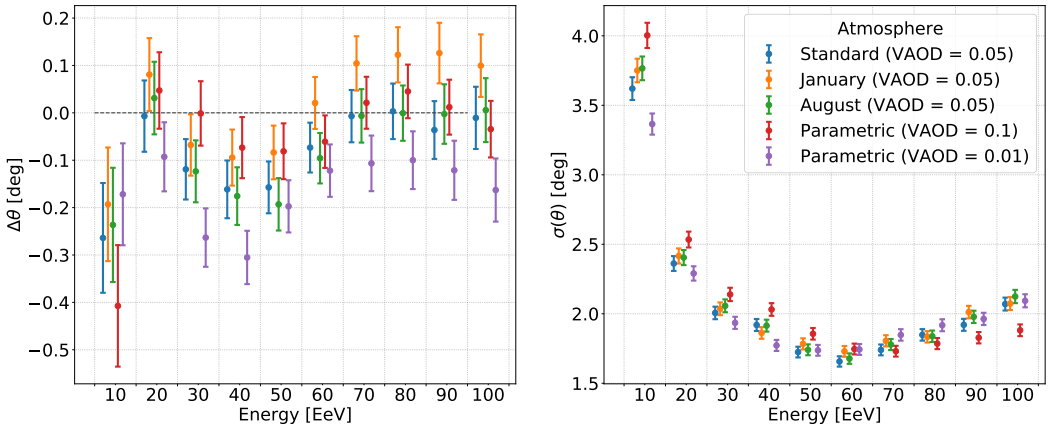


Figure 8.34. The zenith angle bias (left) and resolution (right) as a function of energy for the central core position shown in Figure 8.2 with modified atmospheric models. The data points are shifted laterally to aid the reader. The grid lines separate each energy bin.

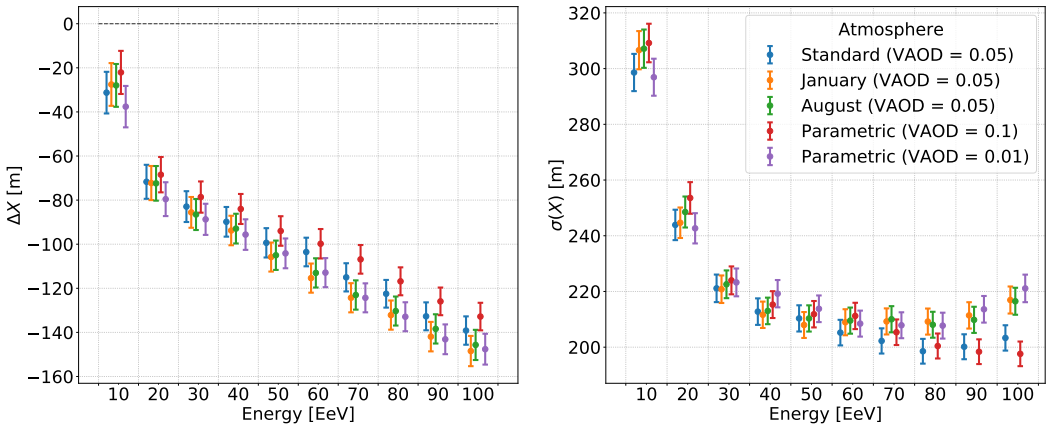


Figure 8.35. The x bias (left) and resolution (right) as a function of energy for the central core position shown in Figure 8.2 with modified atmospheric models. The data points are shifted laterally to aid the reader. The grid lines separate each energy bin.

core position since variations in the y component are more significant than the x component due to the positions of the FAST stations.

8.8 Conclusions

A first guess reconstruction algorithm has been successfully developed for FAST, using a proof-of-concept neural network designed to predict estimates of the shower parameters (X_{max}, energy, θ, ϕ, x and y). The first guess estimates of the shower parameters can be fed into the top-down reconstruction developed in Chapter 7. The performance of the neural network reconstruction is very good for the three representative core regions tested, however, the training should be expanded to

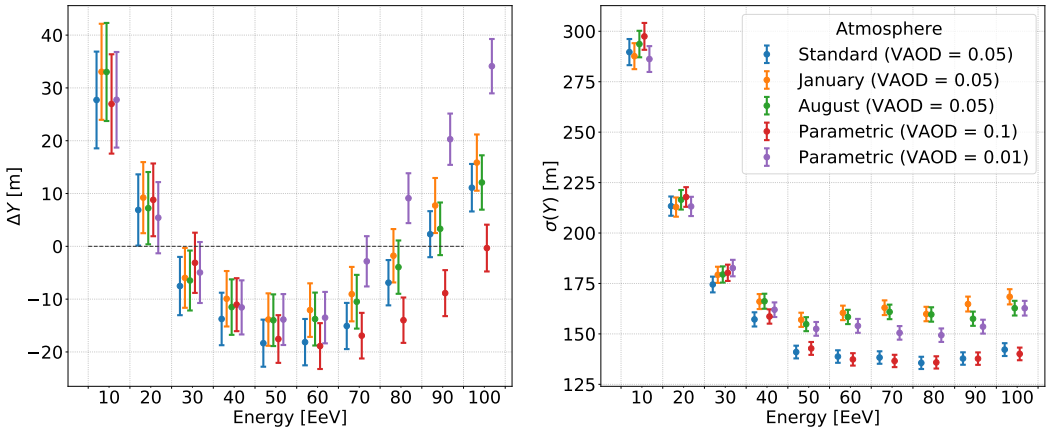


Figure 8.36. The y bias (left) and resolution (right) as a function of energy for the central core position shown in Figure 8.2 with modified atmospheric models. The data points are shifted laterally to aid the reader. The grid lines separate each energy bin.

the full core position phase space for deployment of the reconstruction algorithm on measured data from a future FAST array. The neural network reconstruction is capable of providing a first guess of the shower parameters with an X_{max} resolution of 35-40 g/cm^2, an energy resolution of $\sim$7%, an angular resolution of 2.5-3.5°, and a core resolution of $\sim$250 m at the FAST target energy of $10^{19.5}$ eV.

The developed neural network reconstruction may be capable of acting as a stand-alone reconstruction algorithm with further optimisation of the neural network model and training settings. However, the neural network reconstruction is susceptible to systematic uncertainties since the networks are trained using a fixed detector state. Changes in the atmospheric conditions, for example, can introduce significant reconstruction biases. There is potential to correct for some of these effects, or to train a series of models for varying conditions (e.g. one for each monthly molecular atmosphere model). Future investigation of these issues should be considered in future work.

Given the reliable first guess of the shower parameters provided from the neural network reconstruction developed in this chapter, the top-down reconstruction can provide a final reconstruction of events, taking into account measurements of the atmospheric conditions, calibration and other time-dependent detector effects in the simulation. The combined reconstruction performance of the neural network first guess and the top-down reconstruction for the array reconstruction mode is considered in Chapter 9.

9 FAST Reconstruction Performance

9.1 Introduction

Following the successful development of a first guess reconstruction in Chapter 8 using a neural network, the resultant first guess shower parameters can be passed into the top-down reconstruction developed in Chapter 7. In this chapter, the reconstruction performance of the hybrid and array reconstruction modes are studied. Additionally, several case studies of real event reconstructions using events measured by the the Fluorescence detector Array of Single-pixel Telescopes (FAST) prototype telescopes at the Telescope Array (TA) are considered.

9.2 Top-Down Reconstruction Performance

As discussed in Section 6.1, FAST can operate in two distinct modes. The first of the reconstruction modes is the "hybrid" mode, where a Surface Detector (SD) array is used to independently obtain the shower geometry, leaving only X_{max} and energy to be determined by FAST telescopes. The second mode is the array mode where FAST telescopes operate independently from a ground array. In this mode, the FAST telescopes are arranged in a triangular cell with each station containing 12 telescopes, combining to provide an azimuthal field of view (FoV) of 360°. In this section, the performance of both FAST reconstruction modes are studied in order to determine their viability, as well as their advantages and disadvantages.

9.2.1 Hybrid Mode

As shown in Chapter 7, the top-down reconstruction can be used to fit X_{max} and energy with a fixed geometry obtained from an SD. In order to determine the performance of the hybrid reconstruction mode, simulated events are generated and reconstructed using the top-down reconstruction with the geometry (arrival direction and core position) fixed, leaving only X_{max} and energy free in the likelihood fit. Since an SD reconstruction can only determine the shower geometry within some

resolution, the true arrival directions and core positions of the simulated events are randomly smeared within 1° and 100 m, respectively. Therefore, each event reconstruction uses a fixed geometry different to the true geometry of the event. This accounts for the resolution of the SD geometry reconstruction on average. The reconstruction bias and resolution of X_{max} and energy are studied in the following sections using these simulated events.

9.2.1.1 Geometry Smearing

The geometry of the shower axis as determined by an SD can be described by the reconstructed zenith angle θ, azimuth ϕ, and the x and y components of the core position. These parameters can only be determined within some angular resolution ψ, and core resolution δ. It is important to take this resolution into account when determining the performance of the top-down reconstruction in hybrid mode. While the resolution of a companion SD for FAST is not yet known, the typical resolution of current generation SD arrays is $\sim$1° in arrival direction, and $\sim$100 m in the core position [136].

It is assumed that θ, ϕ, x, and y are all independently normally distributed with mean 0, and variance σ^2_{ang} and σ^2_{core} for the arrival direction and core position parameters, respectively. Both the arrival direction and core position are then distributed as a two-dimensional Gaussian

$$f(z) = \frac{1}{2\pi\sigma^2} z \exp\left(-\frac{z^2}{2\sigma^2}\right) \tag{9.1}$$

where z represents either the space angle (angular distance) α between the true and reconstructed arrival direction, or the ground-distance d, between the true and reconstructed core position, and σ is the appropriate value of either σ_{ang} or σ_{core}. Now the angular resolution ψ, and the core resolution δ are defined as the 68% confidence interval of Equation (9.1). That is, the angular distance containing 68% of reconstructed arrival directions, and the distance containing 68% of reconstructed core positions. Therefore

$$\int_0^{2\pi}\int_0^{\psi} \frac{1}{2\pi\sigma^2} z \exp\left(-\frac{z^2}{2\sigma^2}\right) dz d\phi = 1 - \exp\left(-\frac{\psi^2}{2\sigma^2}\right) = 0.68 \tag{9.2}$$

where the integral over all azimuth ϕ is 2π since the function is azimuthally symmetric. This results in the condition $\psi = 1.51\sigma$. It follows identically that $\delta = 1.51\sigma$. Now given an angular resolution, ψ and core resolution δ, along with Equation (9.1) it is possible to sample a randomly varied arrival direction and core position to emulate the resolution effects of an SD reconstruction.

9.2.1.2 Average Reconstruction Bias and Resolution

In order to determine the FAST hybrid reconstruction performance using the top-down reconstruction, simulated data sets are produced based on the configuration of FAST telescopes shown in Figure 9.1. This configuration matches that of the FAST prototype telescopes at TA. The simulated events are generated at fixed energies of $10^{18.5}$ eV, 10^{19} eV, and $10^{19.5}$ eV with core positions sampled uniformly along lines on the ground with three different azimuth angles ($0°$, $22.5°$, and $37.5°$) up to 15 km from the FAST telescopes, as indicated by the dashed lines in Figure 9.1. The zenith angle of the events is sampled from the standard $\sin\theta\cos\theta$ distribution and the shower azimuth is sampled uniformly, and X_{max} is sampled from a realistic parametrisation of the EPOS-LHC hadronic model as described in Section 8.6. In order to emulate the resolution effects of an SD geometry reconstruction, the true arrival direction and true core position are randomly varied within an angular resolution of $1°$, and a core resolution of 100 m. The FAST prototype telescopes currently operate with an external trigger supplied by the TA Fluorescence Detector (FD). This triggers all three FAST prototype telescopes regardless of the signal. Furthermore, an independent FAST trigger has not yet been developed. Once a proper FAST trigger has been developed, this study should be repeated. Sub-threshold triggers are potentially extremely useful for the top-down reconstruction, but only once an event trigger has been established. Even photomultiplier tubes (PMTs) with no signal are useful in the top-down reconstruction to constrain the shower parameters.

The X_{max} and energy reconstruction bias and resolution are shown as a function of core distance for the $0°$ angle in Figures 9.2 and 9.3, respectively. The X_{max} and energy resolution are best at a core distance of 6-8 km with an X_{max} resolution of $\sim$50 g/cm^2, and an energy resolution of $\sim$20% at $10^{18.5}$ eV. For showers closer to the detector, the resolution becomes significantly worse for both X_{max} and energy. This can be explained by the limited FoV of the FAST telescopes. At a small core distance, the amount of atmospheric depth traversed by the shower inside of the FoV of the telescopes decreases for a given shower geometry, resulting in less of the shower being observed. One solution to this problem would be to increase the elevation FoV of the FAST telescopes to $60°$. An additional effect at small core distances is the angular extent of the spot on the camera. The lateral width of the shower appears wider the closer it is to the detector which effectively reduces the resolution of the camera. At large core distances the resolution improves but begins to degrade for lower energy showers since the signal-to-noise ratio (SNR) of the pulses begins to fall significantly. For the higher energy showers at $10^{19.5}$ eV, the X_{max} resolution remains less than 50 g/cm^2 out to core distances of 15 km, and the energy resolution reaches $\sim$10% at a core distance of 10 km, and remains less than 20% out to core distances of 15 km.

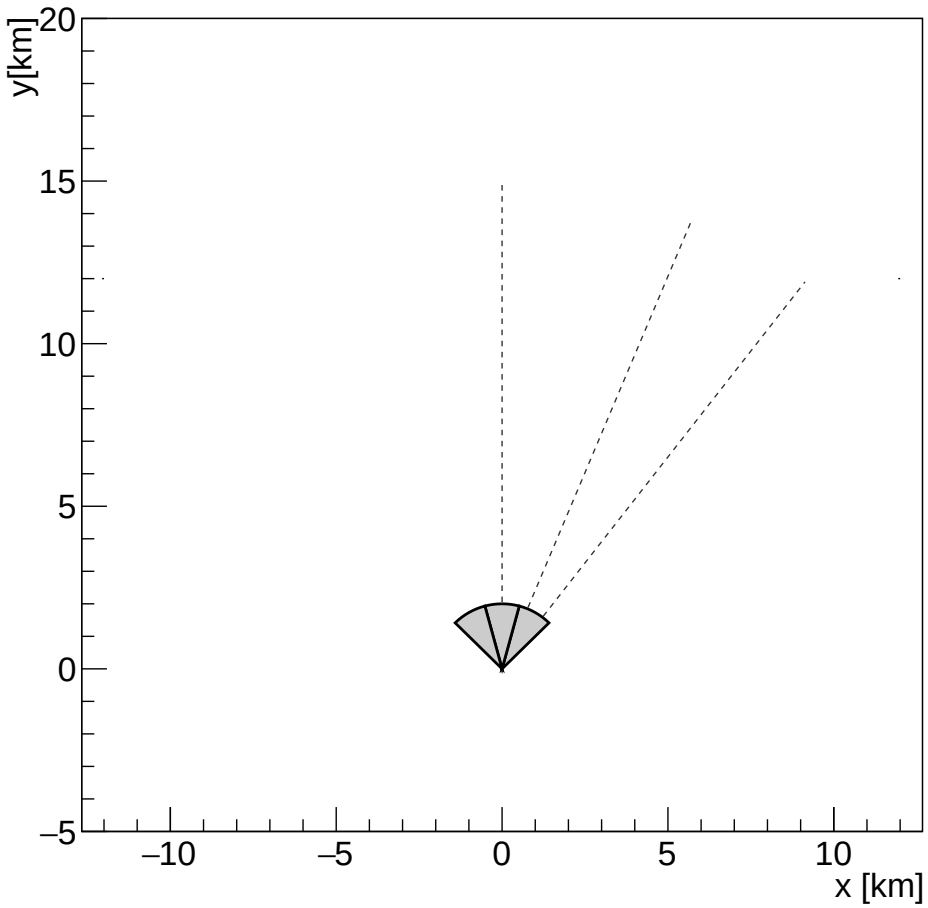


Figure 9.1. The configuration of the simulated FAST telescopes used for the hybrid top-down reconstruction performance. The three dashed lines represent the line of core positions from which events are uniformly sampled. The angles are $0°$, $22.5°$, and $37.5°$.

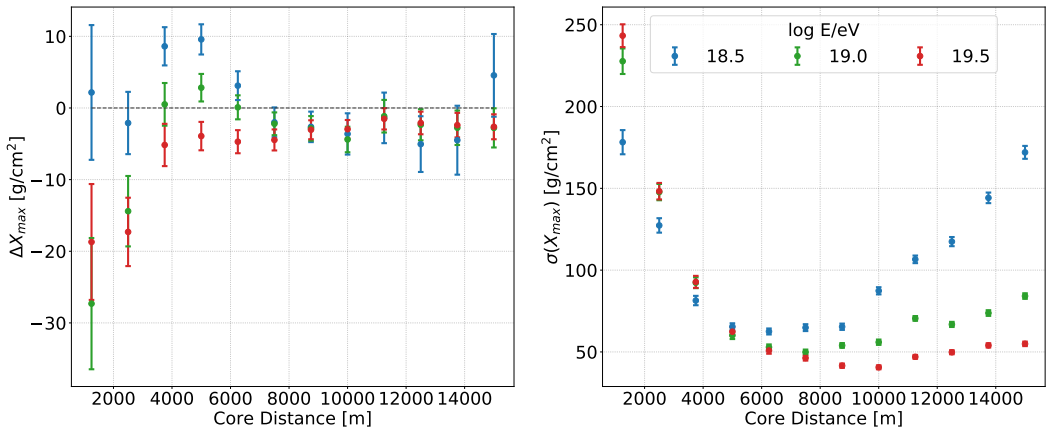


Figure 9.2. The $X_{\max}$ bias (left) and resolution (right) as a function of core distance for three different fixed energies. The events are simulated along a line in the middle of the central FAST telescope shown in Figure 9.1.

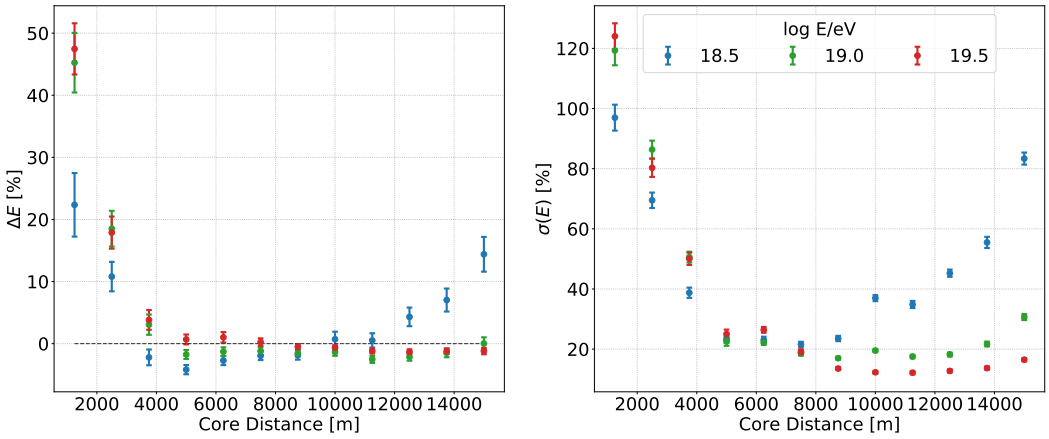


Figure 9.3. The relative energy bias (left) and resolution (right) as a function of core distance for three different fixed energies. The events are simulated along a line in the middle of the central FAST telescope shown in Figure 9.1.

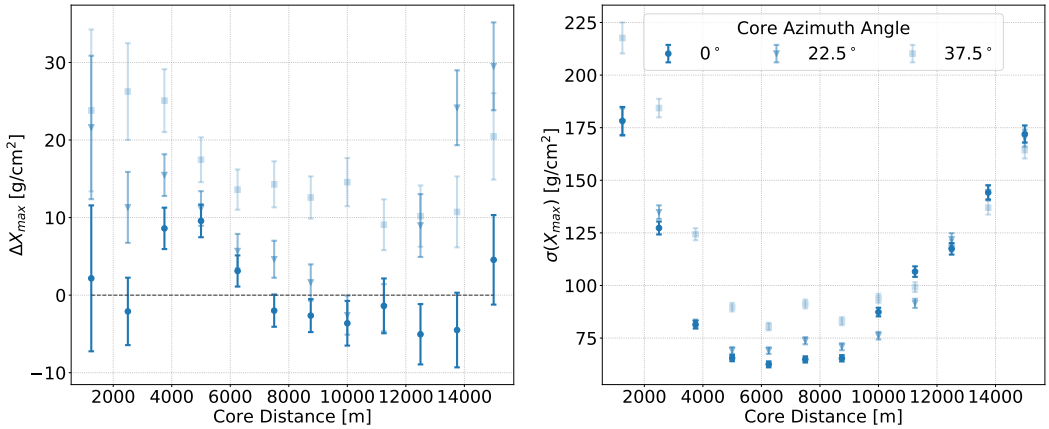


Figure 9.4. The $X_{\max}$ bias (left) and resolution (right) for $10^{18.5}$ eV showers as a function of core distance for three different core azimuth angles represented by the dashed lines shown in Figure 9.1.

In order to test the effect of the FoV, the core positions are simulated along lines at three different angles from the centre of the three telescopes; $0°$, $22.5°$, and $37.5°$. It is expected that showers near the centre of the FoV of the three telescopes will be reconstructed with better resolution on average, while events near the edge of the FoV of the telescopes will be reconstructed with a degraded resolution. The $X_{\max}$ and energy reconstruction bias and resolution are shown as a function of core distance for the three tested angles in Figures 9.4 and 9.5 for $10^{18.5}$ eV showers, and in Figures 9.6 and 9.7 for $10^{19.5}$ eV showers. As expected, the events with core positions closer to the edge of the FoV are generally reconstructed with poorer $X_{\max}$ and energy resolution. The effect is more pronounced at small core distances where the shower is closer to the detector on average. However, at large core distances there is very little difference in the $X_{\max}$ resolution for the events falling near the edge of the FoV.

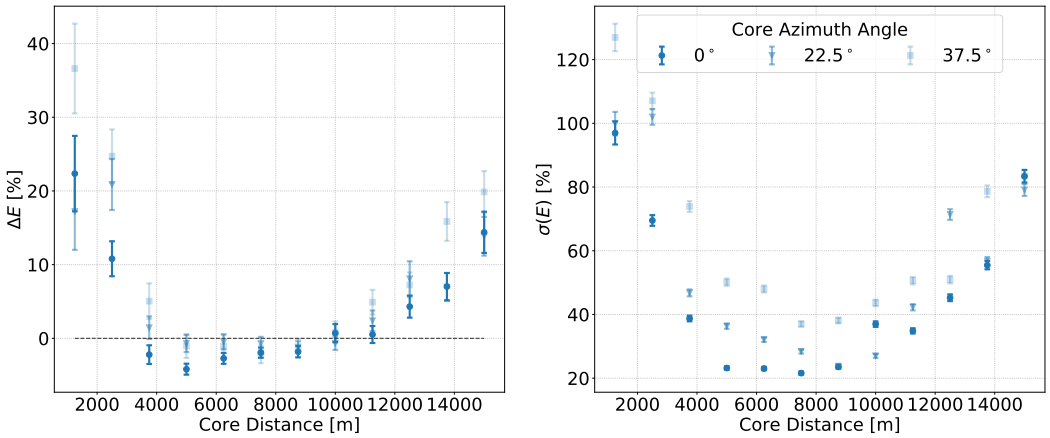


Figure 9.5. The relative energy bias (left) and resolution (right) for $10^{18.5}$ eV showers as a function of core distance for three different core azimuth angles represented by the dashed lines shown in Figure 9.1.

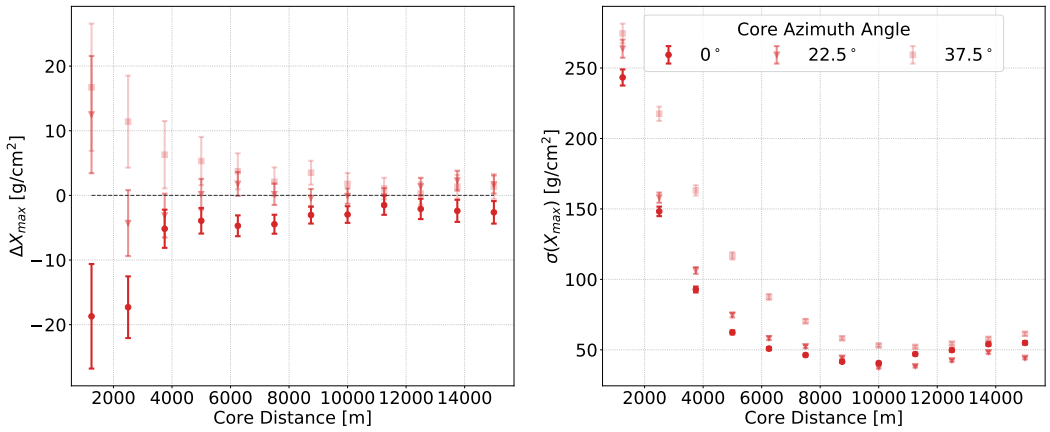


Figure 9.6. The X_{max} bias (left) and resolution (right) for $10^{19.5}$ eV showers as a function of core distance for three different core azimuth angles represented by the dashed lines shown in Figure 9.1.

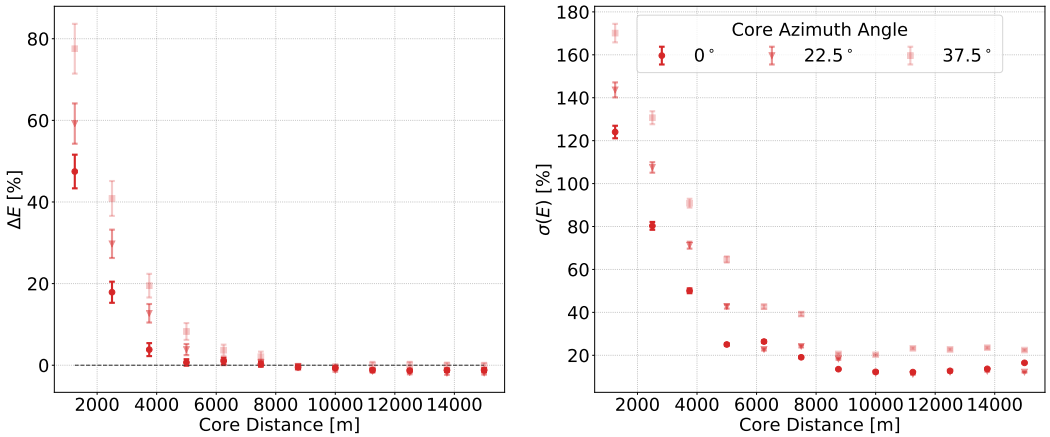


Figure 9.7. The relative energy bias (left) and resolution (right) for $10^{19.5}$ eV showers as a function of core distance for three different core azimuth angles represented by the dashed lines shown in Figure 9.1.

9.2.2 Array Mode

In the array reconstruction mode, the top-down reconstruction is employed to reconstruct all shower parameters simultaneously. This includes X_{max}, energy, zenith, azimuth and the shower core position. The top-down reconstruction alone is not able to converge without a first guess of the shower parameters provided by the neural network reconstruction, which was developed in Chapter 8. In this section the full reconstruction performance of the array mode is studied using simulated events. These simulated events are pre-processed and reconstructed using the neural network reconstruction algorithm to obtain a first guess of each shower parameter. The first guesses are then passed to the top-down reconstruction which obtains the final reconstructed shower parameters.

Since the neural network developed in Chapter 8 was trained on three particular core regions (see Figure 8.2), the simulated events used to study the performance of the top-down reconstruction must be generated within these core regions in order to obtain a reliable estimate of the shower parameters from the trained neural networks. Therefore, the same procedure described in Chapter 8 is used to sample random events. The performance of the top-down reconstruction is discussed in the following sections.

9.2.2.1 Example Reconstruction of a Simulated Event

This section will consider a case study of a full reconstruction where the first guesses of the shower parameters are obtained from the neural network reconstruction, and subsequently passed into the top-down reconstruction. A visualisation of the shower geometry with respect to the three FAST stations is shown in Figure 9.8. The event was generated at an energy of $10^{19.5}$ eV, with an X_{max} value of 725.5 g/cm^2 and a zenith angle of 15.8°. The core position of the event was randomly sampled in the central core region shown in Figure 8.2.

The event was processed through the neural network reconstruction and the first guess of the shower parameters was subsequently passed into the top-down reconstruction to obtain the final reconstructed parameters. The simulated, first guess and final reconstructed values of the shower parameters are summarised in Table 9.1. In this case the first guess of the shower parameters was extremely good, allowing the top-down reconstruction to obtain accurate values of all the shower parameters. This example event illustrates the viability of the developed reconstruction chain including the neural network first guess reconstruction, and the subsequent top-down reconstruction in the ideal case.

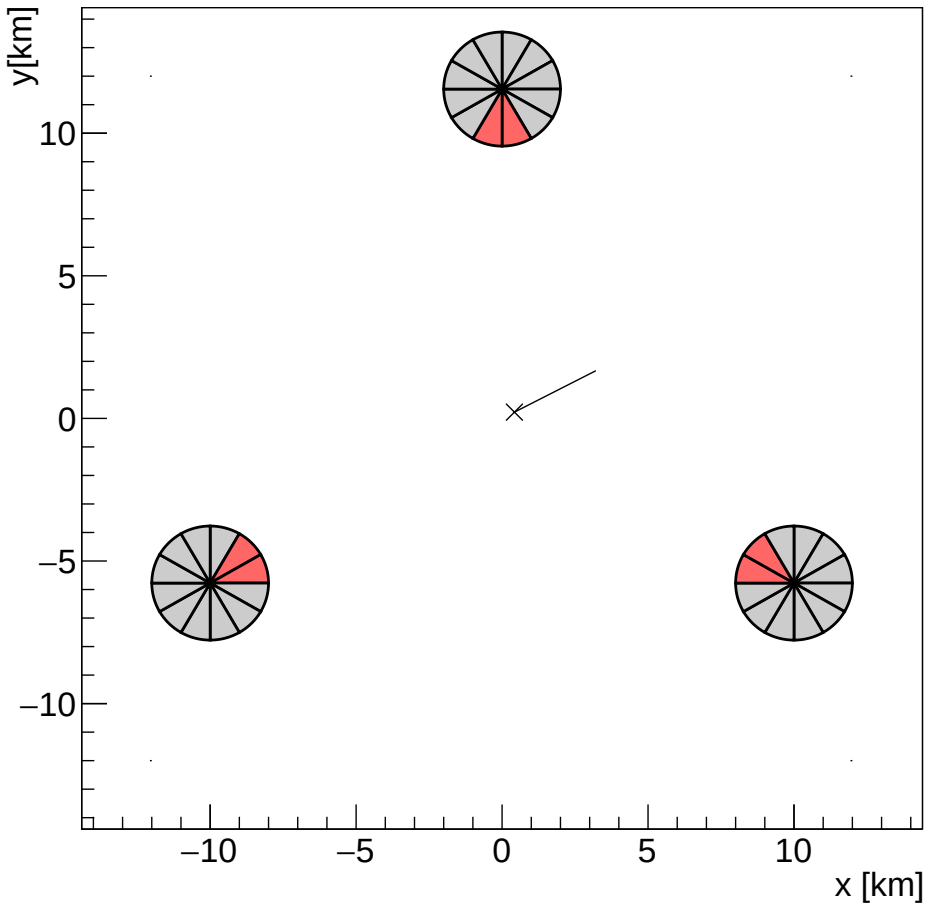


Figure 9.8. The configuration of the FAST telescopes for an example simulated event. The black cross represents the core position of the simulated shower, and the connected black line represents the projection of the shower axis onto the ground. The FAST telescopes highlighted in red represent those which detected signal as determined by the simulated noise-free PMT traces.

Parameter	Simulated Value	First Guess	Reconstructed Value
$X_{\max}$ [g/cm^2]	725.5	711.6	730.4±7.3
Energy [EeV]	31.6	31.9	31.67±0.72
Zenith [deg]	15.8	16.2	15.5±0.2
Azimuth [deg]	27.5	31.0	26.8±1.1
CoreX [m]	423.0	259.4	375.3±33.9
CoreY [m]	219.9	113.6	192.1±27.9

Table 9.1. Summary of the simulated and reconstructed shower parameters for an example simulated event where the event was well reconstructed. The first guess parameters are determined by the neural network reconstruction.

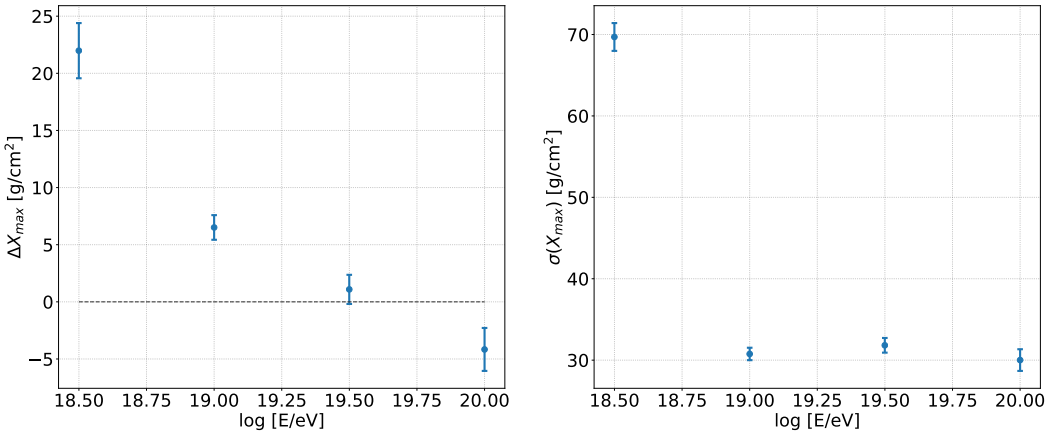


Figure 9.9. The X_{max} bias (left) and resolution (right) as a function of energy for the central core position shown in Figure 8.2.

9.2.2.2 Average Reconstruction Bias and Resolution

Discussed in this section are the average reconstruction bias and resolution of the top-down reconstruction for the central core region shown in Figure 8.2.

The X_{max} reconstruction bias is shown in Figure 9.9(left). At low energy the bias increases to as much as $\sim$22 g/cm^2 at $10^{18.5}$ eV, however, at 10^{19} eV and above, the bias reduces to $\sim$5 g/cm^2. This is similar to the X_{max} reconstruction bias obtained from the neural network first guess reconstruction in Section 8.6. The reconstruction bias at low energy can be explained by a reduction in the trigger efficiency for three stations below 10^{19} eV; at $10^{18.5}$ eV a significant fraction of the events are triggered by only one or two stations, as shown in Section 8.6.5.1. The X_{max} resolution is shown in Figure 9.9(right). The X_{max} resolution is $\sim$30 g/cm^2 at 10^{19} eV and above, however, the resolution significantly degrades to $\sim$70 g/cm^2 at $10^{18.5}$ eV. As shown in Figure 9.10, the X_{max} resolution with the top-down reconstruction is consistent with the resolution of $\sim$30 g/cm^2 at $10^{19.5}$ eV obtained in a previous simulation study performed using a standard Gaisser-Hillas profile fit [215] (see Section 6.3). However, the top-down reconstruction has the (significant) added benefit of a simultaneous geometry reconstruction, rather than a fixed geometry obtained from an SD.

The relative energy bias and resolution are shown in Figure 9.11. The reconstruction bias is almost negligible at less than 0.6%, however, there appears to be a small energy dependence. This dependence indicates that the energy of lower energy showers is slightly underestimated, and the energy of higher energy events is slightly overestimated. The relative energy bias of the neural network first guess in Section 8.6 was small, but larger than the bias from the top-down reconstruction. The energy resolution of the top-down reconstruction is $\sim$7% above 10^{19} eV, however,

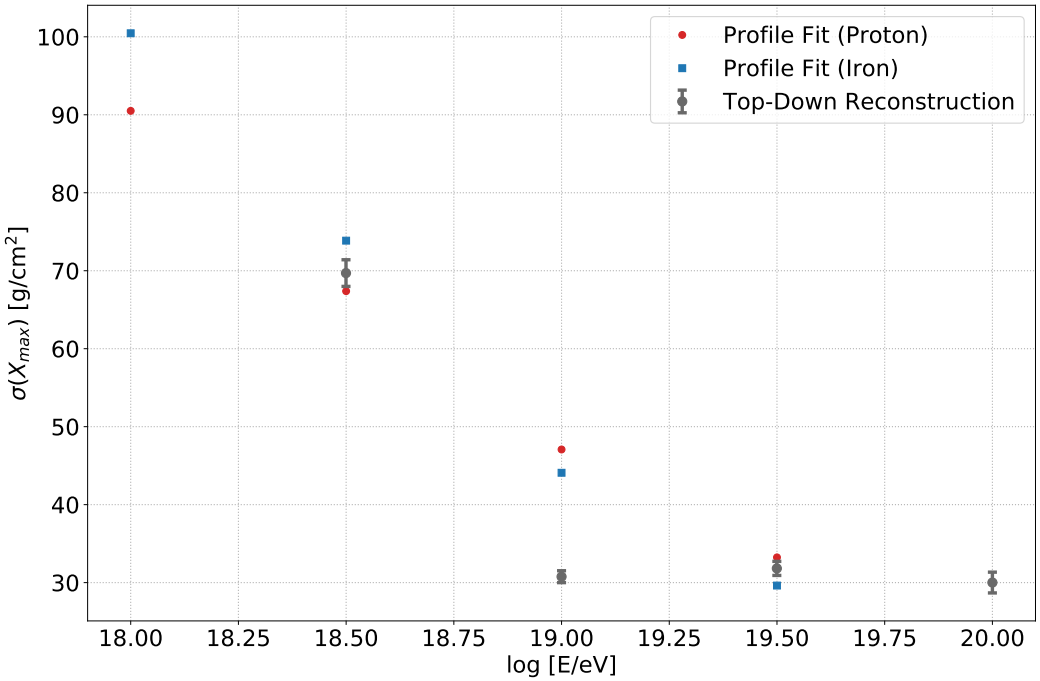


Figure 9.10. The $X_{\max}$ resolution as a function of energy for the central core position shown in Figure 8.2 compared with the results from a previous simulation study performed using a standard Gaisser-Hillas profile fit [215] (see Section 6.3).

it reaches 18% at $10^{18.5}$ eV. Similar energy resolution results were obtained from the neural network first guess in Section 8.6. As for $X_{\max}$, the energy resolution is consistent with the previous simulation study of a reconstruction of the shower profile which obtained a resolution of ∼10% at $10^{19.5}$ eV using a Gaisser-Hillas fit, and a fixed geometry obtained from an SD. This comparison is shown in Figure 9.12.

The zenith angle reconstruction bias is shown in Figure 9.13(left). Similarly to $X_{\max}$ and energy, the zenith angle bias is very small ($\leq$0.2°) at 10^{19} eV and above, however, the zenith angle bias increases to ∼1.3° at 10^{18} eV. The zenith angle resolution, shown in Figure 9.13(right), becomes progressively better with energy, reaching ∼0.75° above the FAST target energy of $10^{19.5}$ eV.

The reconstruction bias and resolution of the x and y components of the core position are shown in Figure 9.14 and Figure 9.15, respectively. The reconstruction bias has a strong energy dependence for both components. However, the bias is relatively small in magnitude: ∼65 m at its most extreme for the x component of the core position at $10^{18.5}$ eV. The resolution is very similar for both components of the core position, reaching 100-150 m at the highest energies, while the resolution is significantly degraded to ∼350 m at $10^{18.5}$ eV, as is the case for the other shower parameters.

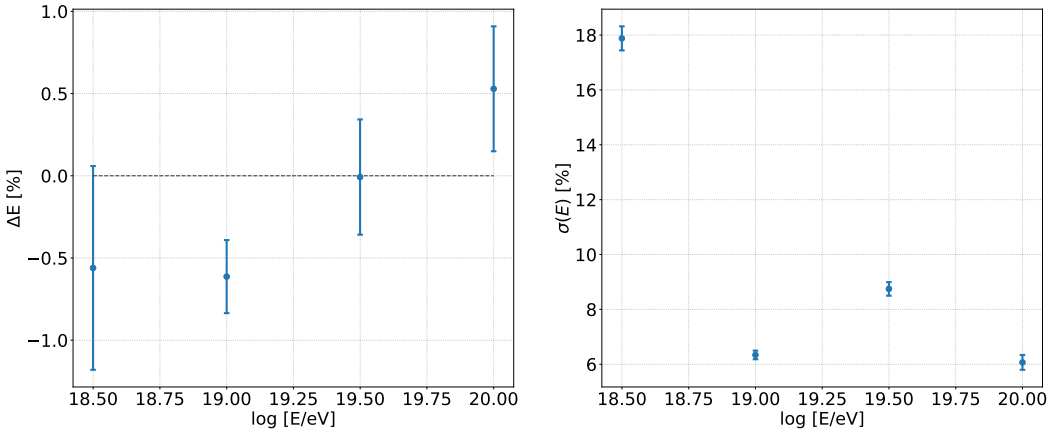


Figure 9.11. The relative energy bias (left) and resolution (right) as a function of energy for the central core position shown in Figure 8.2.

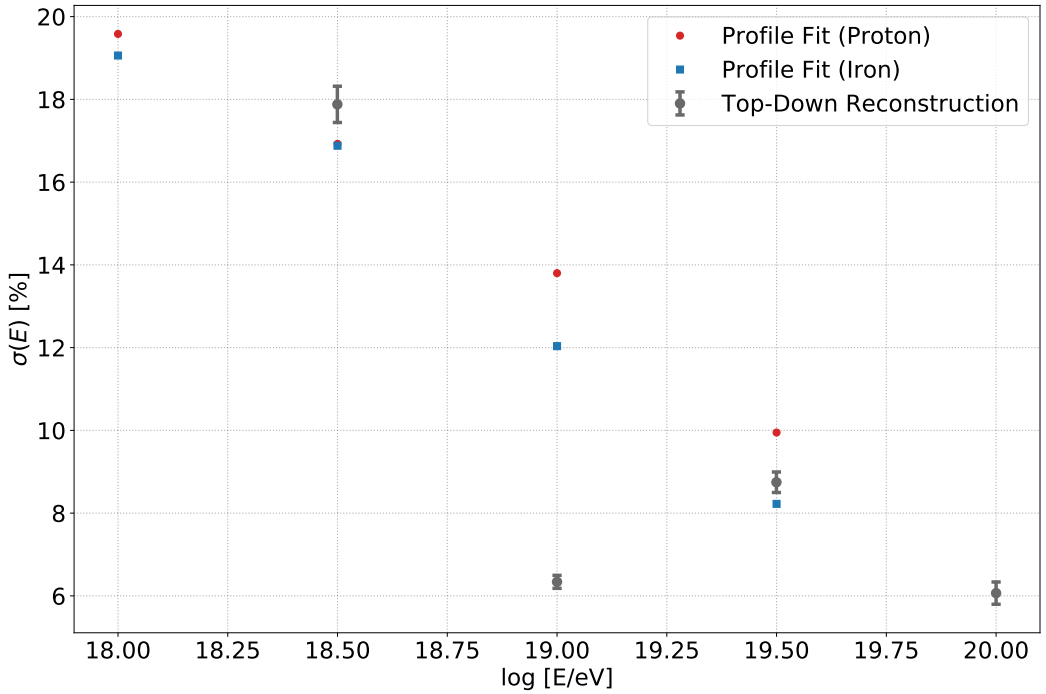


Figure 9.12. The energy resolution as a function of energy for the central core position shown in Figure 8.2 compared with the results from a previous simulation study performed using a standard Gaisser-Hillas profile fit [215] (see Section 6.3).

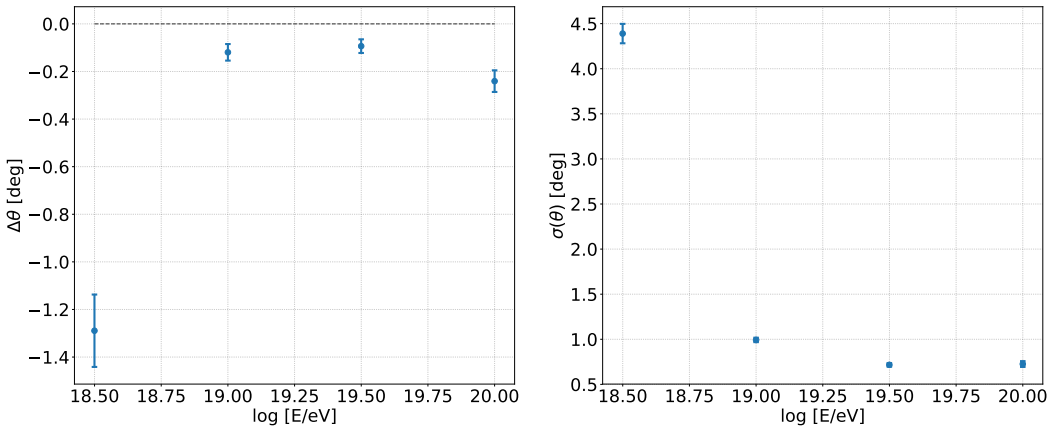


Figure 9.13. The zenith angle bias (left) and resolution (right) as a function of energy for the central core position shown in Figure 8.2.

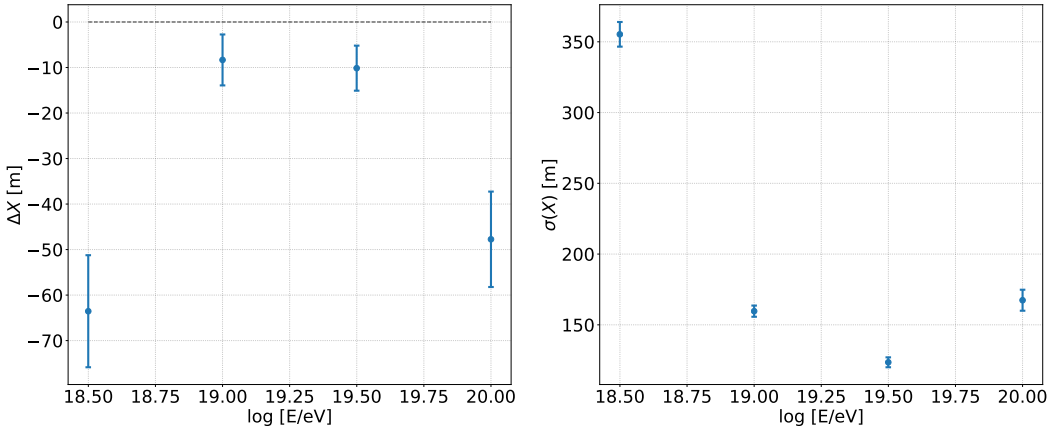


Figure 9.14. The bias (left) and resolution (right) of the x component of the core position as a function of energy for the central core position shown in Figure 8.2.

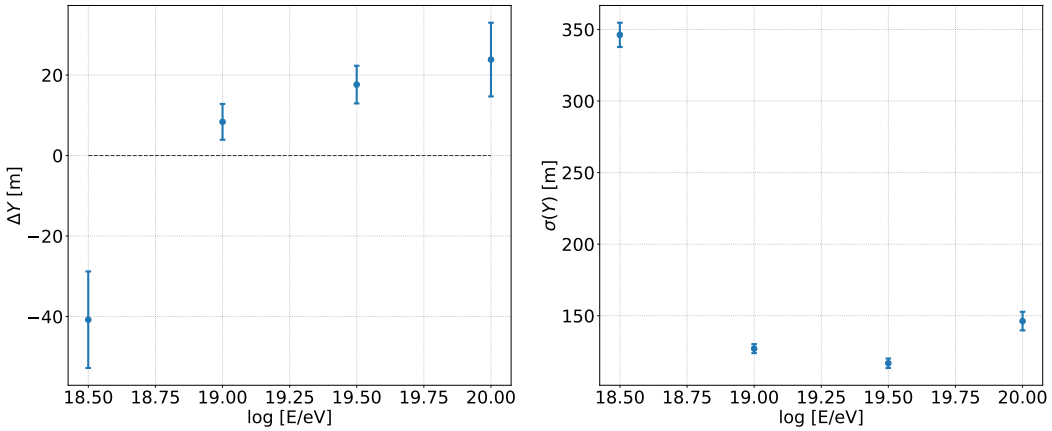


Figure 9.15. The bias (left) and resolution (right) of the y component of the core position as a function of energy for the central core position shown in Figure 8.2.

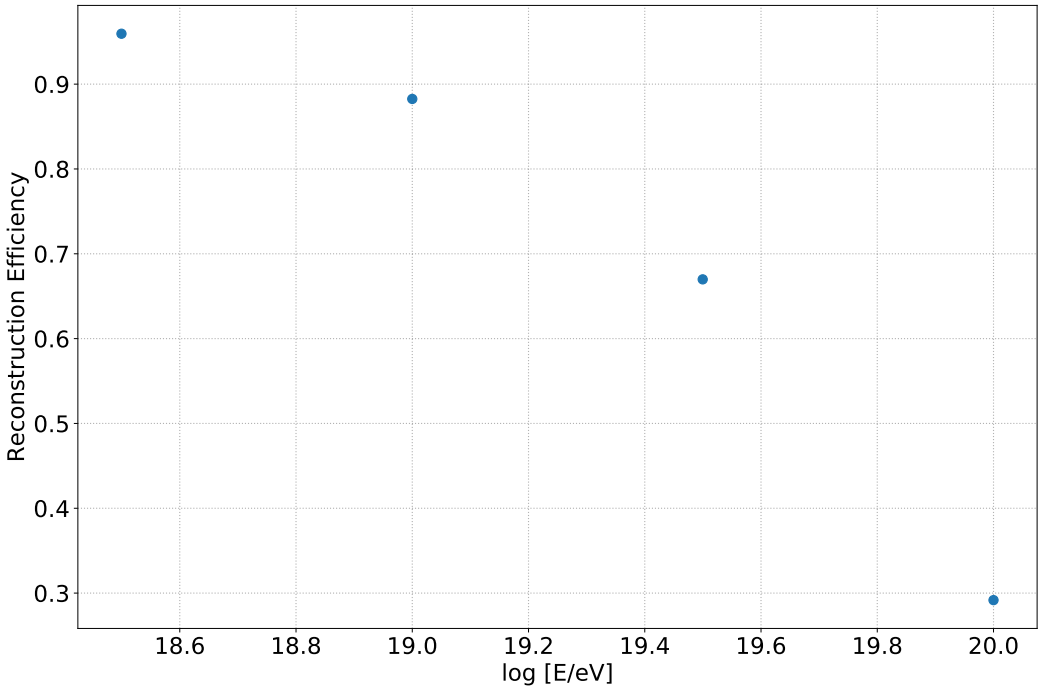


Figure 9.16. The top-down reconstruction efficiency as a function of energy for the central core position shown in Figure 8.2.

The top-down reconstruction efficiency is defined as

$$\epsilon_i = \frac{N_i^{\text{rec}}}{N_i^{\text{trig}}} \tag{9.3}$$

where N_i^{rec} and N_i^{trig} the number of events successfully reconstructed and the number of triggered events, respectively, in energy bin i. The trigger efficiency is discussed in Section 8.6.5.1. The top-down reconstruction efficiency is shown as a function of energy in Figure 9.16. The reconstruction efficiency is acceptable at energies of 10^{19} eV and below, however, at higher energies the reconstruction efficiency sharply decreases. This is a significant issue for FAST since its target energy of $10^{19.5}$ eV is only reconstructed with a $\sim$67% efficiency, and it only gets worse with energy. This opposes expectations since the SNR of the PMT pulses becomes significantly larger at high energy, therefore providing stronger signals. This issue is investigated further in Section 9.2.2.3.

9.2.2.3 Investigating the Efficiency Problem

The inefficiency at $10^{19.5}$ eV and above can be explained by inaccuracies in the detector simulation. Higher energy events, where the pulses have a larger SNR, are more susceptible to this issue. The simulated signals are calculated in discrete time bins, and very small steps in the shower parameters will only change the

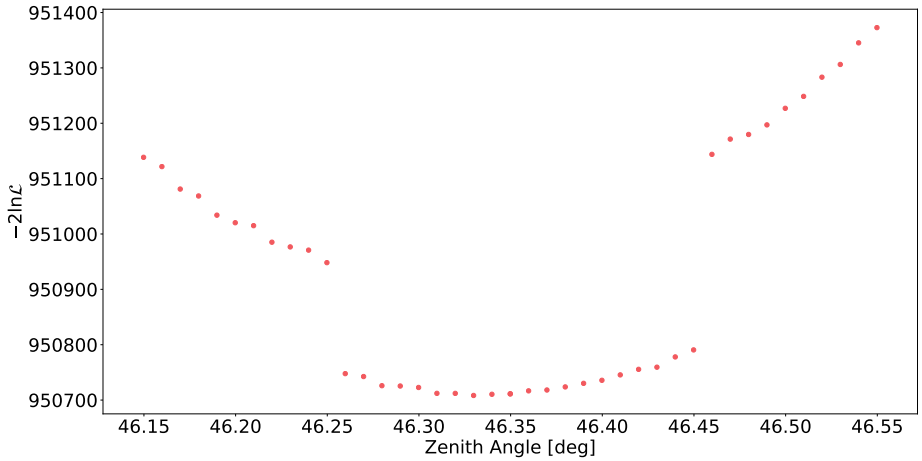


Figure 9.17. An example scan of the likelihood function as a function of zenith angle with all other shower parameters fixed. The energy of this simulated event is 10^{20} eV.

signals on the scale of these time bins. This leads to inaccuracies in the simulation, especially for high energy events. Lower energy events have smaller SNR and therefore larger parameter uncertainties. Therefore, the minimiser does not attempt to search the likelihood function at such small steps of the parameter space. This can be understood with an example event, for which the minimiser fails to converge on a minimum. The example event was simulated at an energy of 10^{20} eV, an X_{max} value of 783.6 g/cm^2, and a zenith angle of 46.3°. To illustrate the inaccuracy of the simulation at very small scales, the value of the likelihood function is scanned over zenith angle with all other shower parameters fixed to their true values. The result of the scan over zenith angle is shown in Figure 9.17. The scan over zenith angle highlights the inaccuracy of the simulation at very small scales. The large jumps in the likelihood function are a significant issue for most standard minimisers which usually rely on accurate estimates of the derivative of the likelihood function (e.g. gradient descent). With larger parameter uncertainties, the minimiser searches the parameter space on a significantly larger scale, where the likelihood function more closely approximates a smooth function. This is illustrated in Figure 9.18, which shows the same scan of the likelihood function over zenith angle shown in Figure 9.17, but over a large range, with larger steps in zenith angle. Note that the jumps in the likelihood are still present, but they are less prevalent at the larger scale.

In order to account for this issue, the simulation should be optimised to produce a smooth likelihood function, even at these extremely high energies. With default settings, the detector simulation calculates the signal in time bins of 10 ns, before re-binning the signals into 100 ns bins. However, it is possible to increase the time bin resolution by calculating the signal in smaller time bins (e.g. 1 ns time

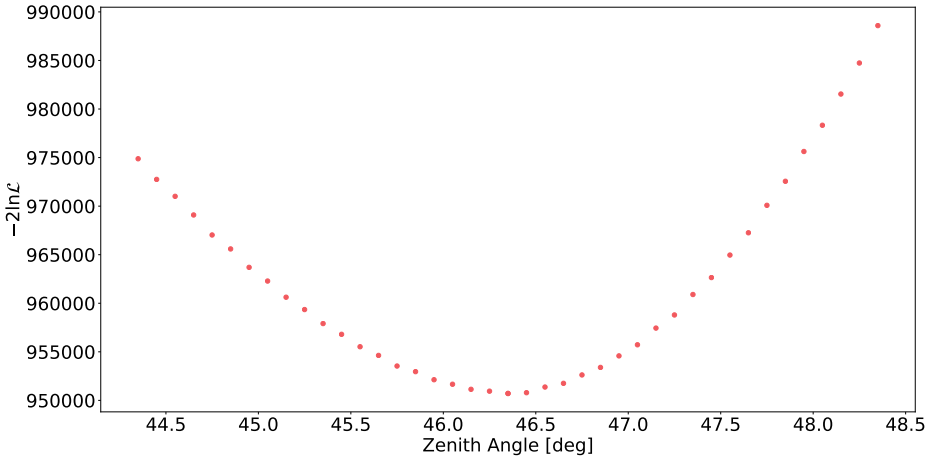


Figure 9.18. The example scan of the likelihood function as a function of zenith angle from Figure 9.17, but scanned over a larger scale, with smaller steps in zenith angle.

bins). The disadvantage of this approach is that the simulation requires 10 times the processing time. This is impractical for the already computationally expensive nature of the top-down reconstruction. Another approach is to circumvent this issue by artificially reducing the value of the likelihood function. This has the effect of forcing the minimiser to search the parameter space in larger steps, since a given comparison between measured and simulated events will yield a smaller value of the likelihood function, and events with varied shower parameters will therefore appear to match better than they would have without the scaling. Hence, more extreme shower parameters will appear acceptable to the minimiser according to the scaled value of the likelihood function. Recall that the parameter uncertainties are determined by the increase of the likelihood function by a value determined by the number of degrees of freedom in the likelihood fit, as described in Section 7.2.6. For the six shower parameters freely fit in the array reconstruction mode, the value of $-2\ln\mathcal{L}$ must increase by ∼7.04. So, if the value of the likelihood artificially increases at a faster rate, then the inaccuracies are less important. However, this will cause parameter uncertainties to be overestimated. The top-down reconstruction likelihood function developed in Chapter 7 has the form

$$-2\ln\mathcal{L}(\vec{x}\,|\vec{a}\,) = \sum_{i}^{N} \frac{(x_i - A\mu_i)^2}{[\sigma^2 + A\mu_i\,(1+V_g)]} + \ln\left[\sigma^2 + A\mu_i\,(1+V_g)\right] + \ln 2\pi \qquad (9.4)$$

where x_i and μ_i are the signals of the measured and simulated event, respectively, in time bin i at a reference energy of 10^{19} eV, and A is the energy scale factor. The approach of reducing the magnitude of the likelihood function to force the minimiser to search the parameter space on a larger scale can be achieved by increasing the variance of the simulated background noise. This is not unphysical since operation

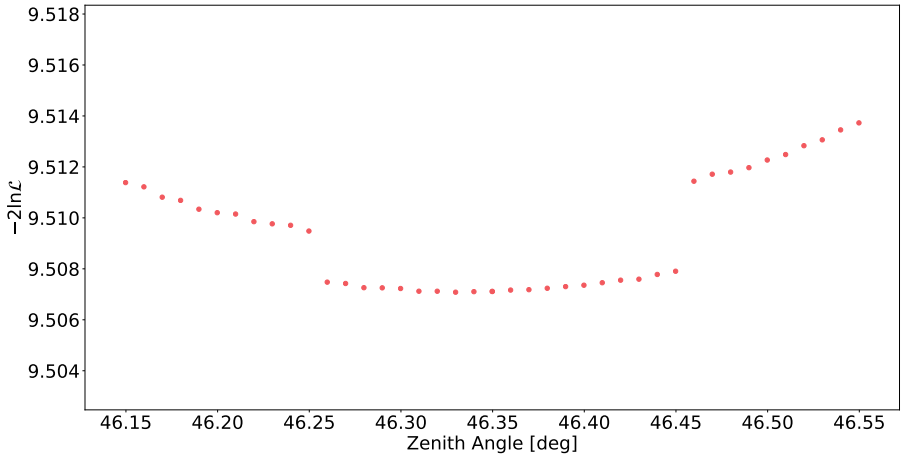


Figure 9.19. The example scan of the likelihood function as a function of zenith angle from Figure 9.17, but with the likelihood function scaled by a factor of 10^{-5}.

under moonlight for example, significantly increases the night-sky background (NSB). However, it is not necessary to modify the actual background noise of the simulated events. Instead, the output of the likelihood function can be directly scaled to achieve the same effect. The scaled likelihood function is simply given by

$$\ln \mathcal{L}_{\mathrm{S}} = S \ln \mathcal{L} \tag{9.5}$$

where $\ln \mathcal{L}_{\mathrm{S}}$ is the scaled, and $\ln \mathcal{L}$ is the unscaled log-likelihood, and S is the likelihood scale factor. In order to increase the uncertainty and reduce the change in the likelihood function for a given step size in the multidimensional parameter space, a value of $S < 0$ must be chosen. To illustrate the effect of the scaled likelihood function, the previous scan over zenith angle is repeated in Figure 9.19. Note that the jumps in the likelihood function are still present, however, the scale of these jumps is much smaller than the parameter uncertainties, which correspond to an increase of 1 in $-2 \ln \mathcal{L}$ for a single parameter, or $\sim$7.04 for all six shower parameters.

In order to test the likelihood scaling approach, and to further illustrate the efficiency problem, the previous performance tests are repeated with the scaled likelihood function used during the minimisation with a scale factor of 10^{-5}. The reconstruction bias and resolution of the shower parameters as obtained for the top-down reconstruction with a scaled likelihood function are compared with the results from the unscaled likelihood function in Figures 9.20 to 9.26. The reconstruction bias in each parameter is relatively unaffected by the scaling of the likelihood function, the only exception being at low energy, where the bias appears to increase slightly in each case. The resolution of all shower parameters is also degraded. This is a consequence of scaling the likelihood function. Since the minimiser searched

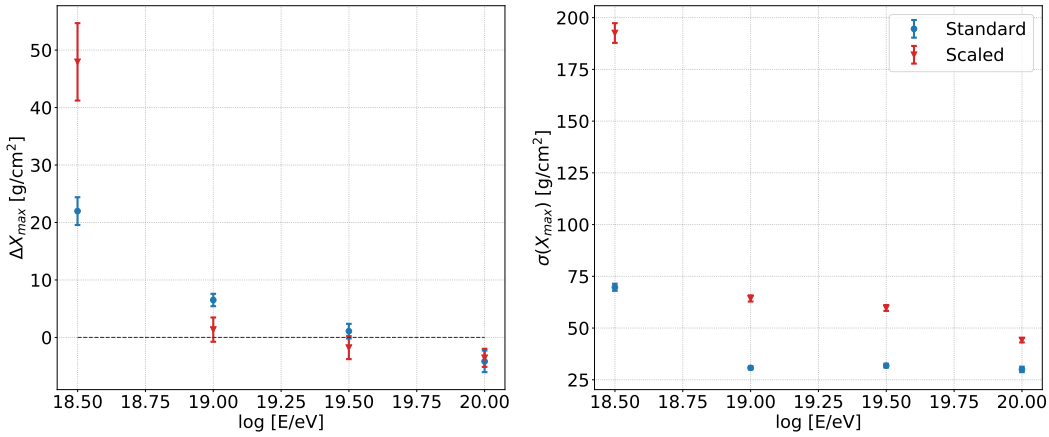


Figure 9.20. The $X_{\max}$ bias (left) and resolution (right) as a function of energy for the central core position shown in Figure 8.2.

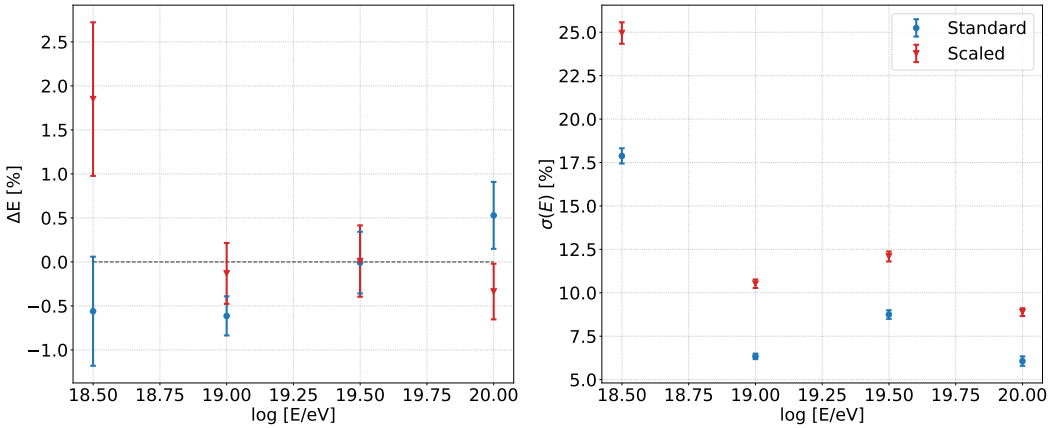


Figure 9.21. The relative energy bias (left) and resolution (right) as a function of energy for the central core position shown in Figure 8.2.

the parameter space on a significantly larger scale, and the value of the likelihood function is smaller in magnitude, the requirements for convergence on a minimum are more relaxed relative to the unscaled likelihood function. This means that the top-down reconstruction generally converges on a minimum with shower parameters which are further from the true shower parameters compared to the case of the unscaled likelihood function.

The reconstruction efficiency using the scaled likelihood function is also compared to the unscaled likelihood function in Figure 9.27. A significant improvement in the reconstruction efficiency is evident. This gives further credence to the idea that the imprecision of the simulation directly causes the efficiency problem. The problem is not completely resolved, however. Further investigation into the issues in the detector simulation should be considered in future work. It may be possible to improve the detector simulation so that the jumps in the likelihood function are removed. However, if this is not possible, then further effort to improve the robustness of the

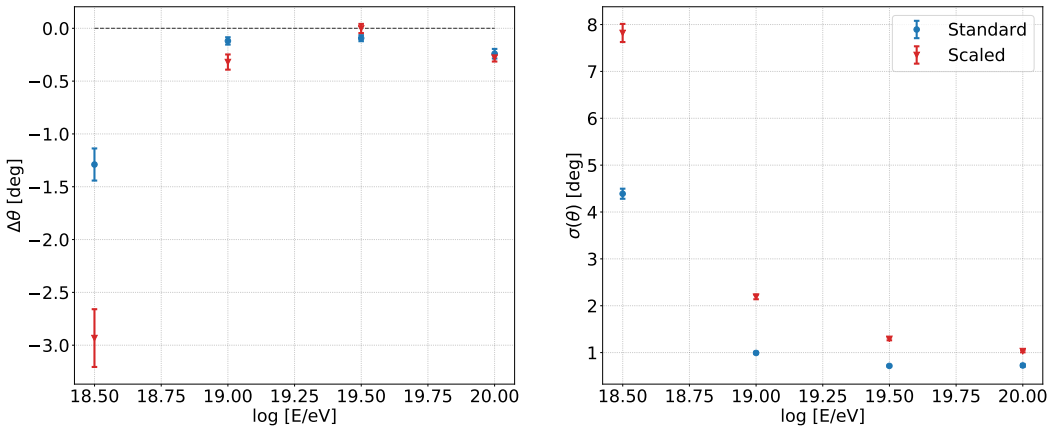


Figure 9.22. The zenith angle bias (left) and resolution (right) as a function of energy for the central core position shown in Figure 8.2.

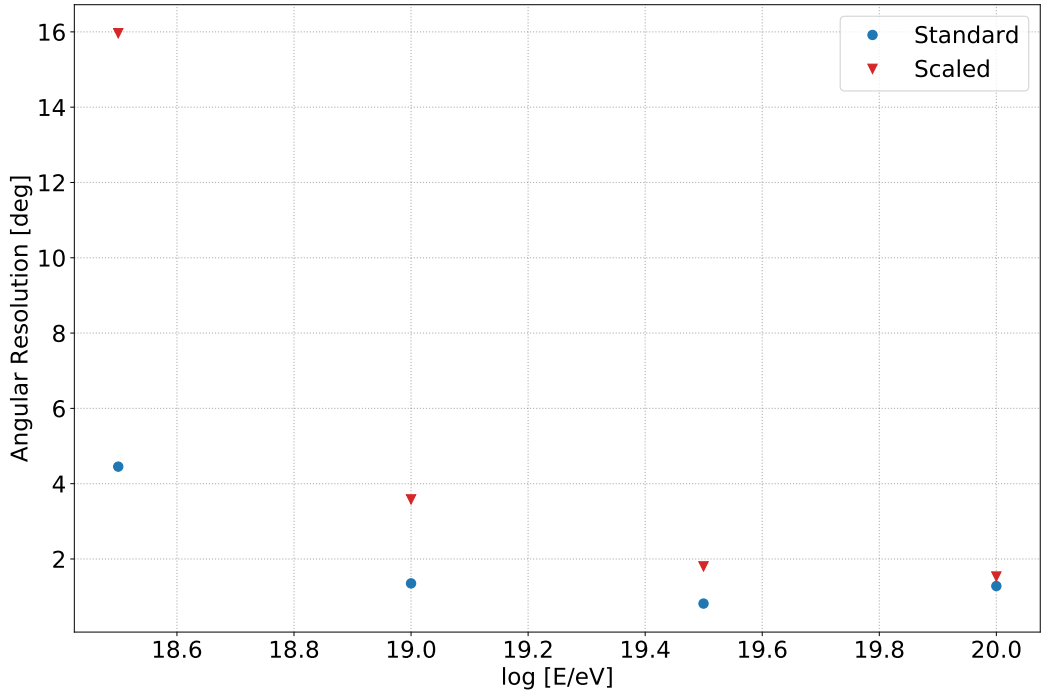


Figure 9.23. The angular resolution as a function of energy for the central core position shown in Figure 8.2.

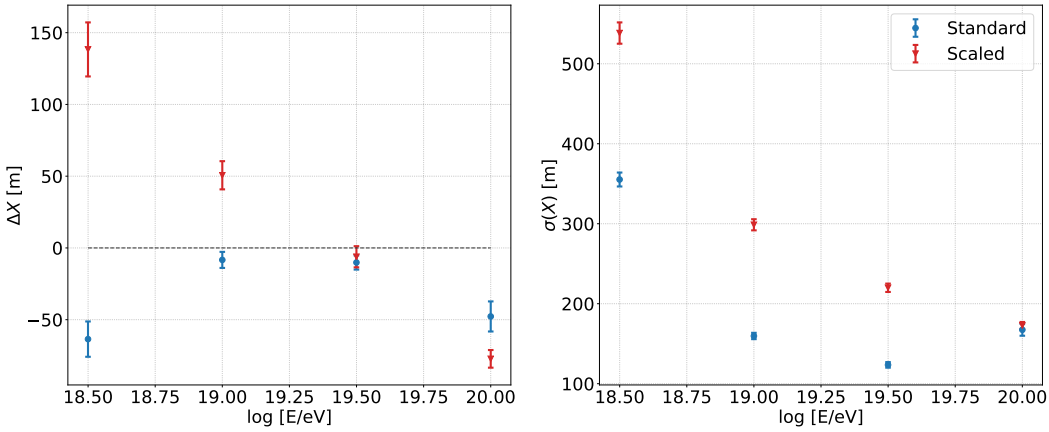


Figure 9.24. The bias (left) and resolution (right) of the x component of the core position as a function of energy for the central core position shown in Figure 8.2.

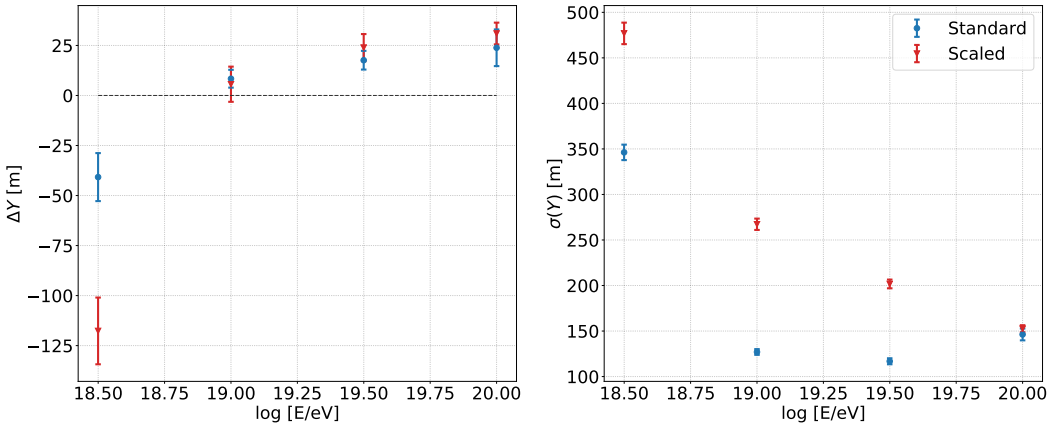


Figure 9.25. The bias (left) and resolution (right) of the y component of the core position as a function of energy for the central core position shown in Figure 8.2.

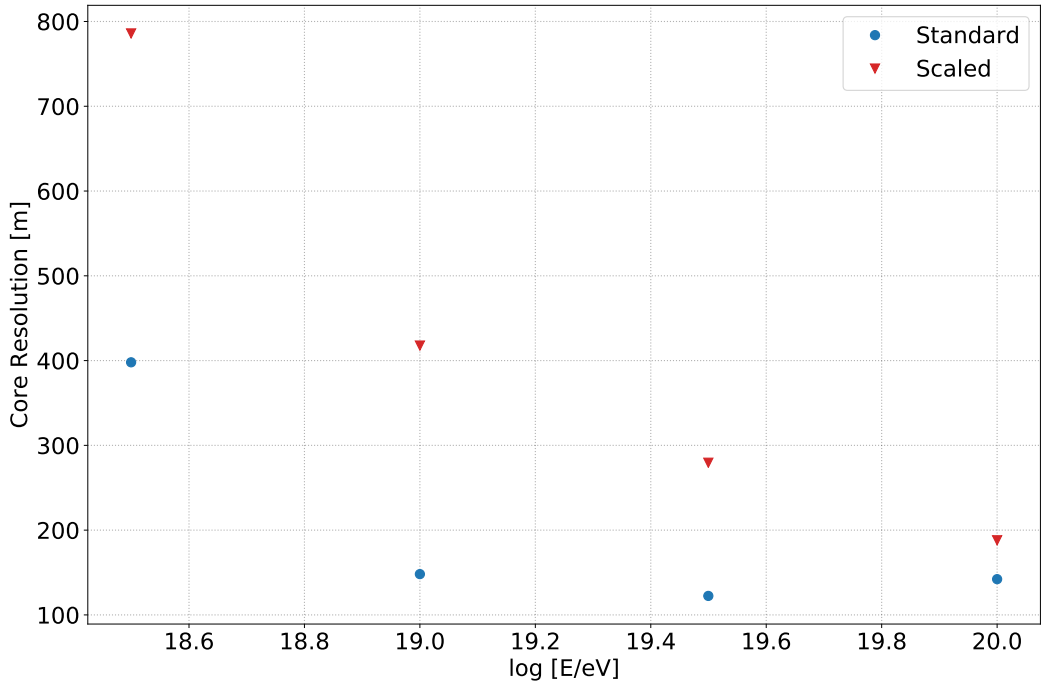


Figure 9.26. The core resolution as a function of energy for the central core position shown in Figure 8.2.

minimiser should be considered. The cost of this improved reconstruction efficiency is reduced resolution in the reconstructed shower parameters. Nevertheless, the top-down reconstruction with the scaled likelihood function still produces adequate resolution in the reconstructed shower parameters, with the possibility of future improvements to the detector simulation or the minimisation procedure. The top-down reconstruction using the scaled (unscaled) likelihood function obtains an X_{max} resolution of $\sim$60 g/cm^2 (30 g/cm^2), an energy resolution of $\sim$12% (8%), an angular resolution of $\sim$1.8° (0.8°) and core resolution of $\sim$280 m (120 m) at the FAST target energy of $10^{19.5}$ eV.

9.2.2.4 Dependence on Core Position

In a similar way to the neural network reconstruction, the performance of the top-down reconstruction has been evaluated for the three core regions shown in Figure 8.2. The top-down reconstruction performance has been evaluated using the scaled likelihood function described in Section 9.2.2.3 in order to maximise the reconstruction efficiency of the simulated events, and to determine a realistic reconstruction performance given the efficiency issues discussed previously. The reconstruction bias and resolution for the upper and right core regions shown in Figure 8.2 are compared to that of the central core position from Section 9.2.2.3.

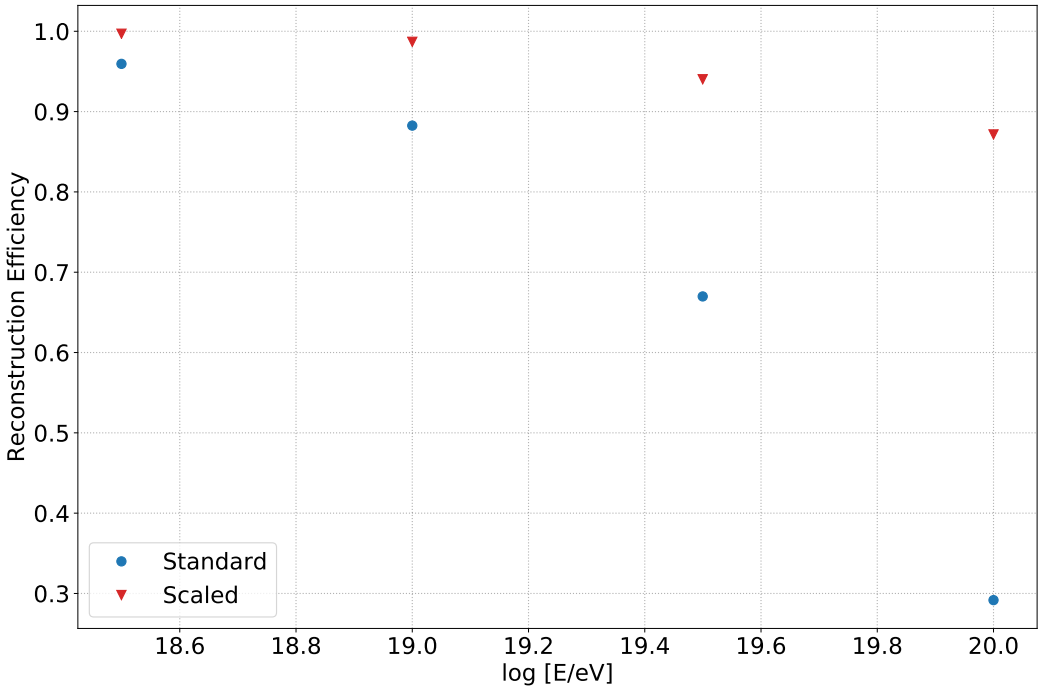


Figure 9.27. The top-down reconstruction efficiency as a function of energy for the central core position shown in Figure 8.2.

The X_{max} reconstruction bias and resolution are shown in Figure 9.28 for the three core positions. The reconstruction bias shows little core dependence, with the exception of the lowest energy bin of $10^{18.5}$ eV. The X_{max} reconstruction bias is less than 10 g/cm^2 for all core positions above 10^{19} eV. The X_{max} resolution also shows little core dependence, except in the lowest energy bin where the resolution in the central core region is significantly worse compared to the other two regions. This is likely due to the reduced efficiency for three stations in the central core position at lower energy compared to the other two core positions, as discussed in Section 8.6.5.1.

The relative energy bias and resolution are shown in Figure 9.29. The upper core position generally performs worse than the other two core positions, however, there is also little core dependence in the reconstruction bias with the exception of the lowest energy bin. The resolution is generally best for the central core position and higher energy, and worse for the upper core position. However, the core dependence becomes smaller with increasing energy. This is likely due to the fact that the trigger efficiency of the FAST stations increases with energy.

As for X_{max} and energy, the zenith angle performance shows little core dependence. However, the reconstruction bias (Figure 9.30(left)) is larger for the central core position, followed by the right and upper core positions in the lowest energy bin. At higher energy, the reconstruction bias is relatively small and independent of

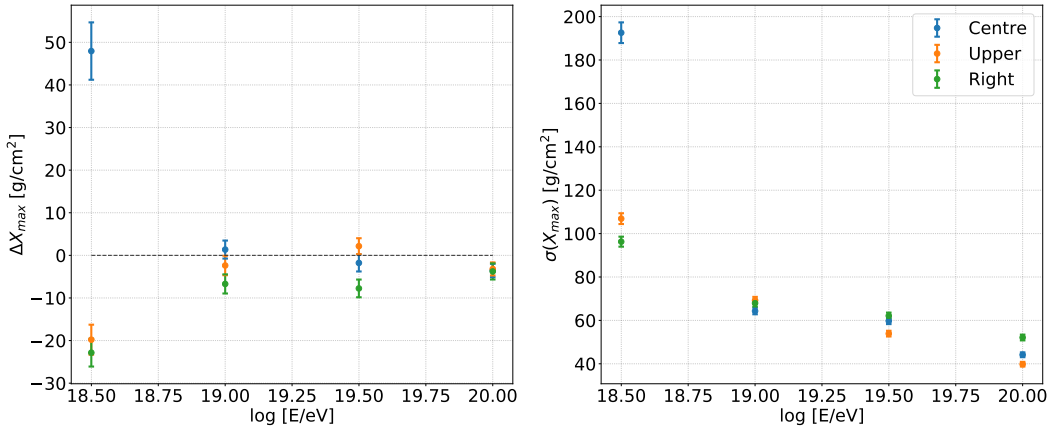


Figure 9.28. The $X_{\max}$ bias (left) and resolution (right) as a function of energy for the three core positions shown in Figure 8.2.

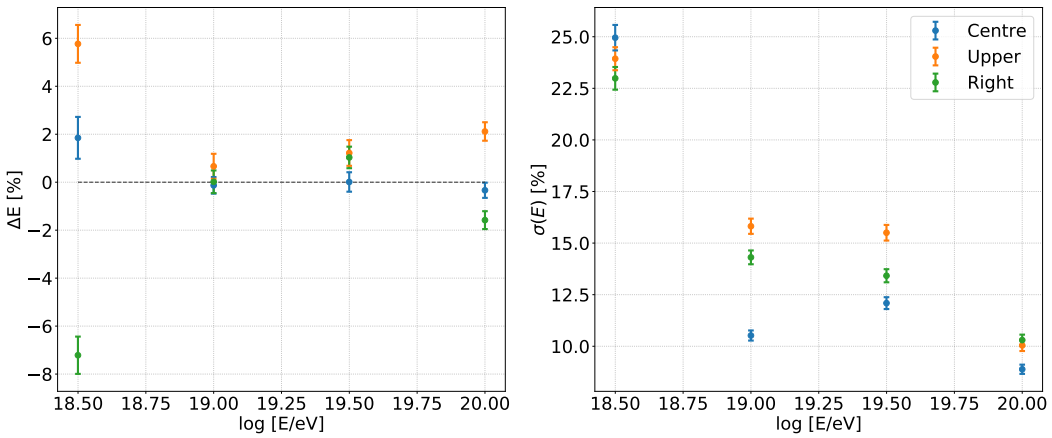


Figure 9.29. The relative energy bias (left) and resolution (right) as a function of energy for the three core positions shown in Figure 8.2.

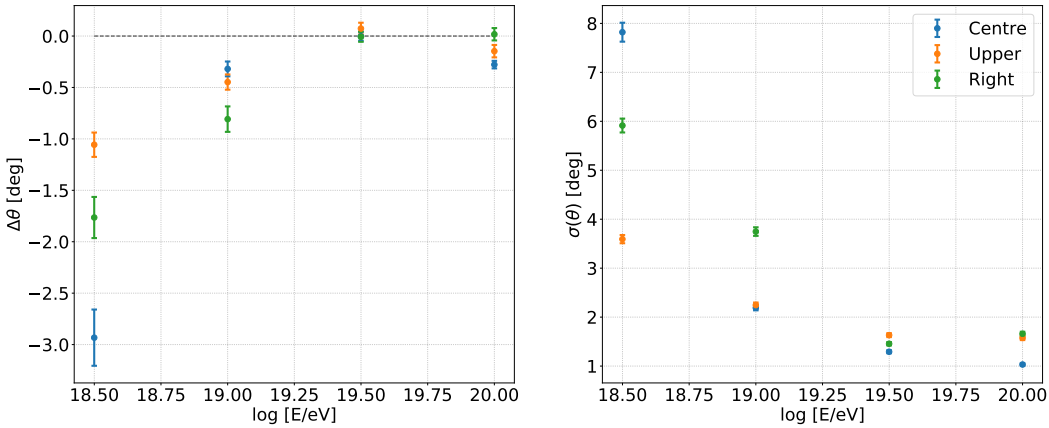


Figure 9.30. The zenith angle bias (left) and resolution (right) as a function of energy for the three core positions shown in Figure 8.2.

core position. The zenith angle resolution, shown in Figure 9.30 (right) also shows very little core dependence. The combined angular resolution of the top-down reconstruction for the three core positions is shown in Figure 9.31. The angular resolution has been determined using the same procedure described in Section 8.6.3. The angular resolution significantly improves with energy, from $\sim$16° for the central core position at $10^{18.5}$ eV, to less than 2° above $10^{19.5}$ eV. At higher energy, the angular resolution of the upper core position is slightly worse than the other two core positions.

The reconstruction bias and resolution of the x and y components of the core position are shown in Figures 9.32 and 9.33, respectively. Some minor core dependence is observed, however, the overall reconstruction bias is small over this energy range. Also shown in Figure 9.34 is the combined core resolution, determined as described in Section 8.6.4. The core resolution reaches 200-300 m at the FAST target energy of $10^{19.5}$ eV.

The reconstruction efficiency of the three core positions is also compared in Figure 9.35. There is very little core dependence on the reconstruction efficiency.

9.2.2.5 Atmospheric Systematics

As discussed in Section 8.7, the neural network reconstruction suffers from significant systematic uncertainties due to the fixed detector state used to simulate events for training. In this section, these systematic uncertainties are revisited to determine if the top-down reconstruction is able to recover the bias in the first guess shower parameters under the assumption that the true state of the detector can be

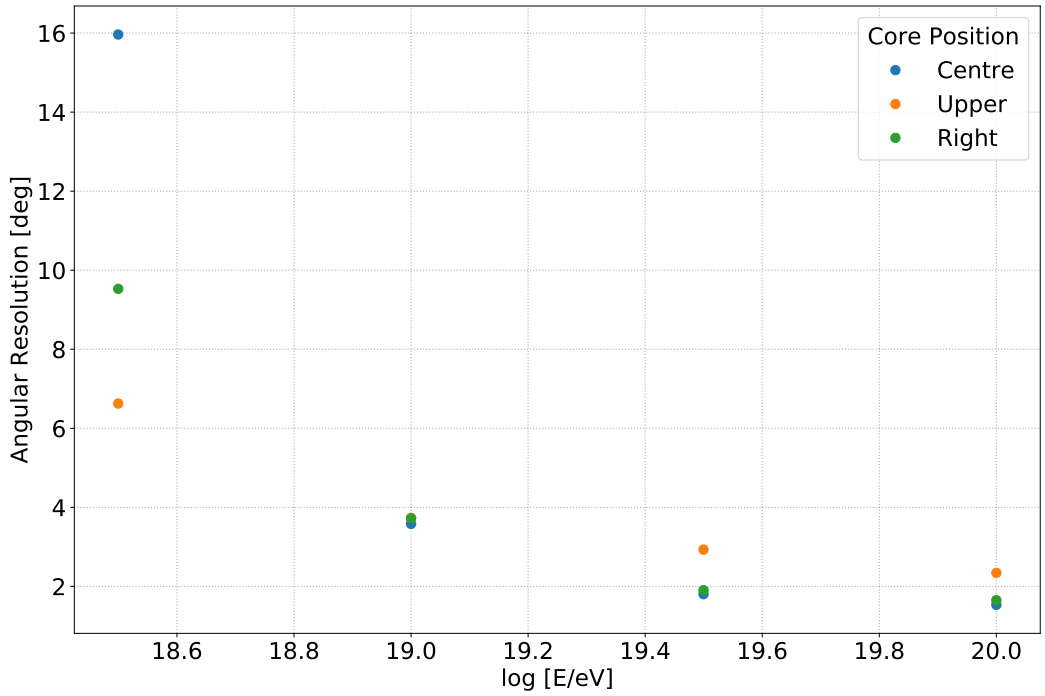


Figure 9.31. The angular resolution as a function of energy for the three core positions shown in Figure 8.2.

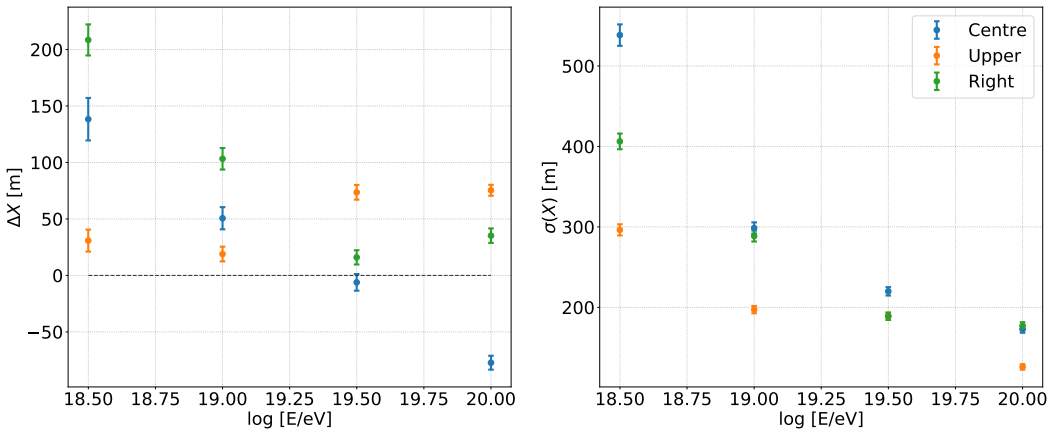


Figure 9.32. The bias (left) and resolution (right) of the x component of the core position as a function of energy for the three core positions shown in Figure 8.2.

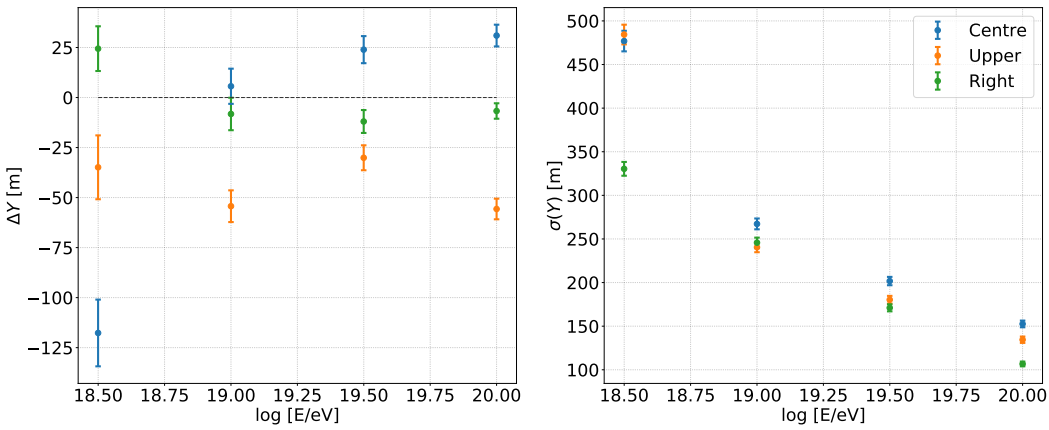


Figure 9.33. The bias (left) and resolution (right) of the y component of the core position as a function of energy for the three core positions shown in Figure 8.2.

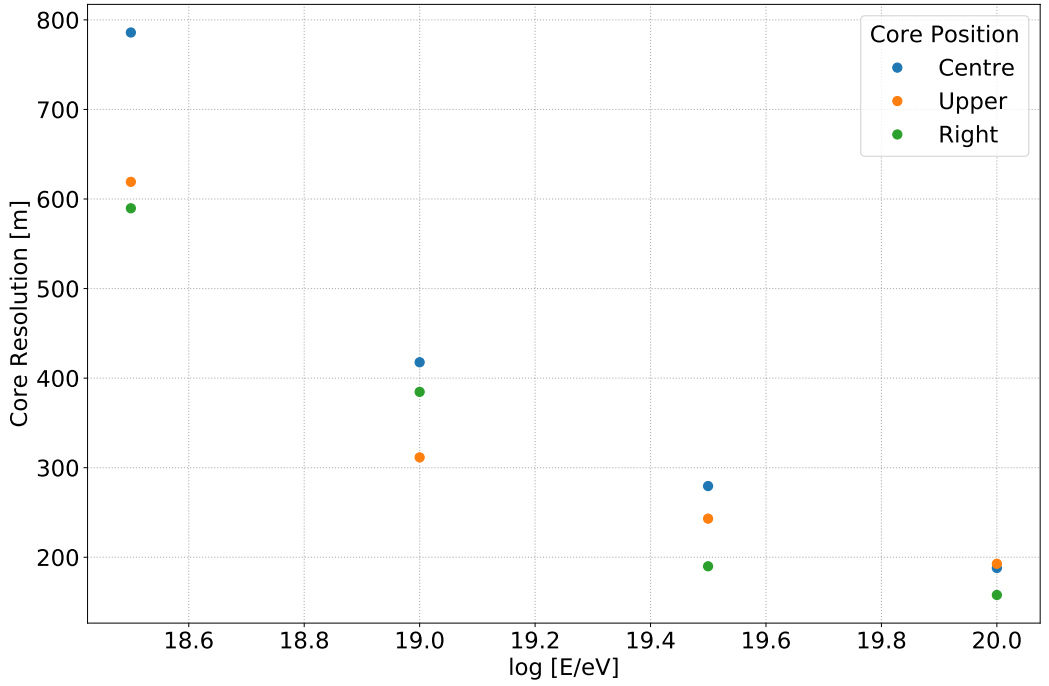


Figure 9.34. The core resolution as a function of energy for the three core positions shown in Figure 8.2.

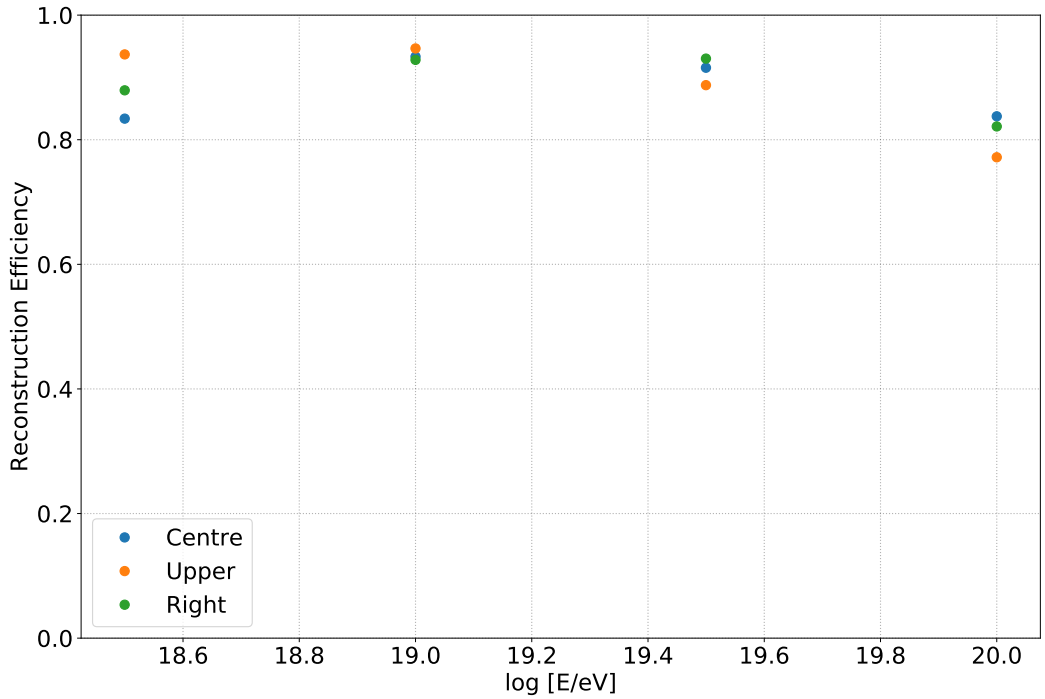


Figure 9.35. The top-down reconstruction efficiency as a function of energy for the three core positions shown in Figure 8.2.

measured and corrected for in the detector simulation employed in the top-down reconstruction.

The tests in the section are focused on the systematic bias caused by alternate atmospheric models compared to the model used to train the neural network. The most extreme bias in the reconstructed X_{max} was produced using the January model of the molecular atmosphere. Therefore, this model was chosen to test the top-down reconstruction's ability to recover a biased first guess. Similarly, the largest bias in the energy reconstruction was caused by changes in the vertical aerosol optical depth (VAOD) of the simulated atmosphere. A second simulated data set was produced with a VAOD of 0.1 compared to the VAOD of $\sim$0.055 which the neural network was trained on. The two simulated data sets were pre-processed, and each of the simulated events was first reconstructed with the neural network trained on simulations using the standard atmosphere. The events with a biased first guess of the shower parameters were then passed into the top-down reconstruction to determine the final estimate of the shower parameters.

The neural network reconstruction produces an X_{max} reconstruction bias of $\sim$30 g/cm^2 above 10^{19} eV for events simulated with the January molecular atmosphere model. This is shown in Figure 9.36(left) along with the reconstruction bias of the subsequent top-down reconstruction. The reconstruction bias caused by the fixed detector state in the simulations used to train the neural network is almost completely re-

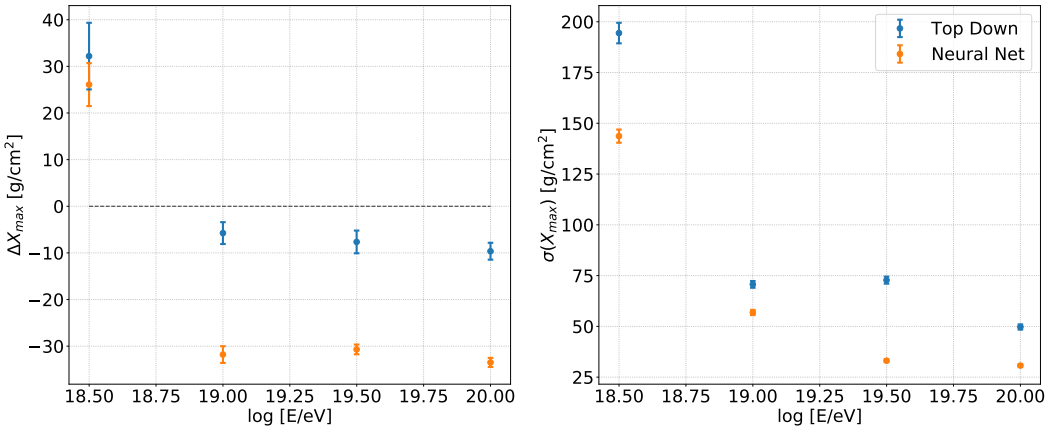


Figure 9.36. The $X_{\max}$ bias (left) and resolution (right) as a function of energy for the top-down reconstruction (blue) and the neural network reconstruction (orange). These events were simulated with the January molecular model instead of the parametric model used to train the neural network.

moved by the top-down reconstruction. The only exception is at lower energy, where the reconstruction bias of the top-down reconstruction is large even in the ideal case as shown in Figure 9.28. The energy reconstruction bias caused by the alternate molecular model is also recovered by the top-down reconstruction, as shown in Figure 9.37. These results further validate the entire reconstruction chain including the neural network as a first guess of the shower parameters, and the top-down reconstruction as the final estimate of the shower parameters. This also suggests that it will not be necessary to account for the reconstruction bias of the neural network when used in conjunction with the top-down reconstruction. The poorer $X_{\max}$ and energy resolution of the top-down reconstruction compared to the neural network reconstruction can be explained by the scaled likelihood function. With improvements to the detector simulation, this scaling will be unnecessary, and the resolution of the top-down reconstruction will improve significantly.

For completeness the reconstruction bias and resolution of the zenith angle, and the x and y components of the core position are also shown in Figures 9.39 and 9.40. As discussed in Section 8.7 these parameters did not show any significant increase in their reconstruction bias due to changes in the molecular atmosphere model. However, it is worth noting that the reconstruction bias of each of these parameters is not significantly different compared to the neural network first guess and the top-down reconstruction. There are small improvements in the zenith angle and core resolution for the top-down reconstruction compared to the neural network at high energy.

In a similar way to the change in the molecular model, the reconstruction bias and resolution for each of the shower parameters was determined for the neural network first guess reconstruction, and the subsequent top-down reconstruction

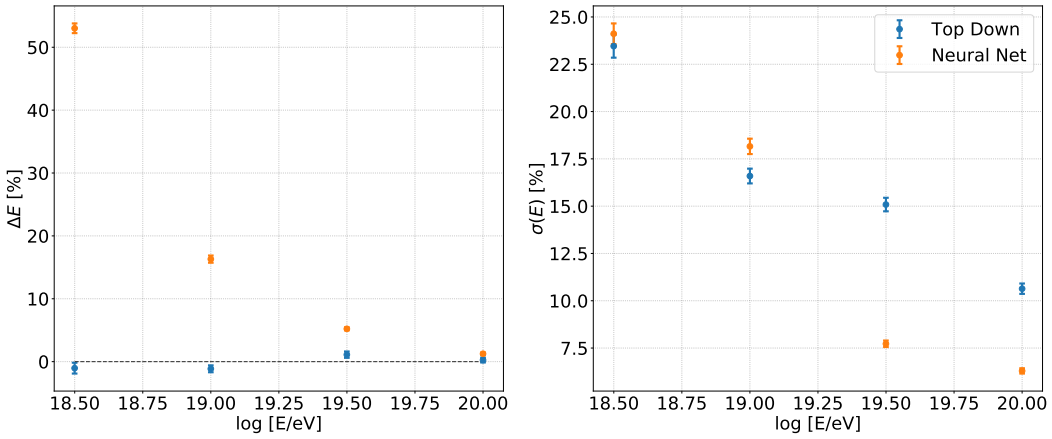


Figure 9.37. The relative energy bias (left) and resolution (right) as a function of energy for the top-down reconstruction (blue) and the neural network reconstruction (orange). These events were simulated with the January molecular model instead of the parametric model used to train the neural network.

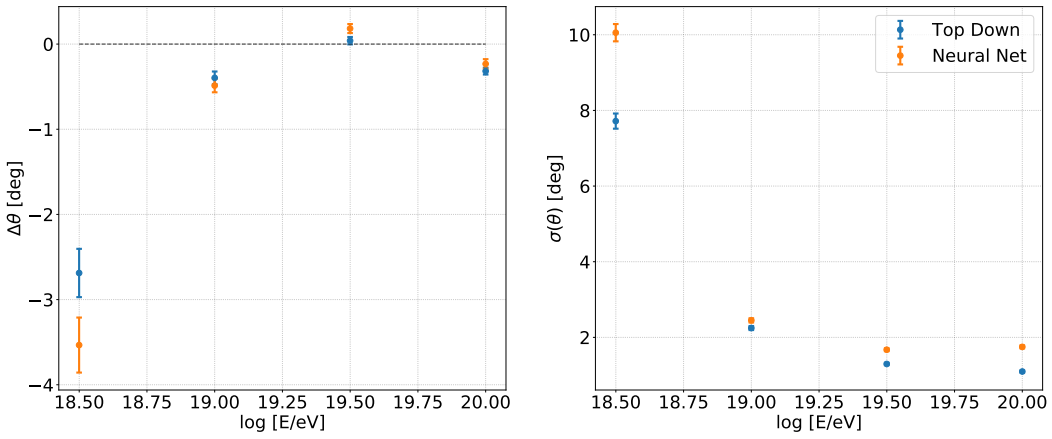


Figure 9.38. The zenith angle bias (left) and resolution (right) as a function of energy for the top-down reconstruction (blue) and the neural network reconstruction (orange). These events were simulated with the January molecular model instead of the parametric model used to train the neural network.

.

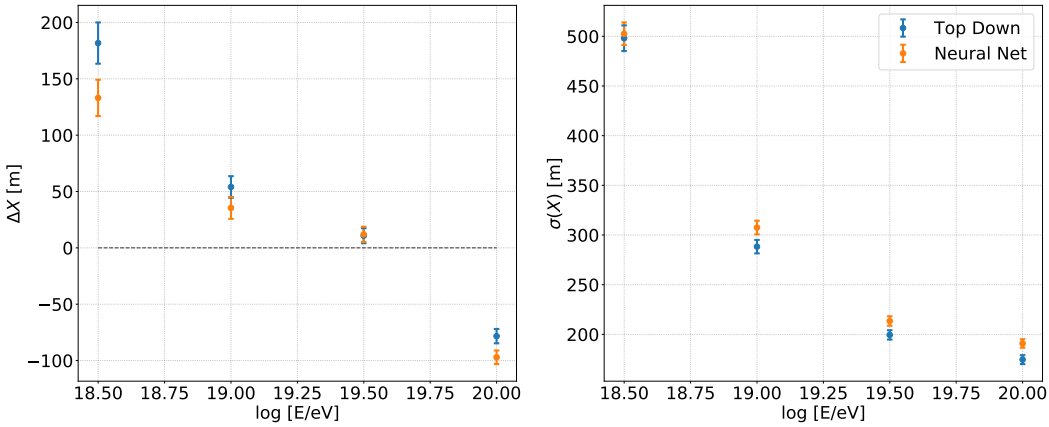


Figure 9.39. The bias (left) and resolution (right) of the x component of the core position as a function of energy for the top-down reconstruction (blue) and the neural network reconstruction (orange). These events were simulated with the January molecular model instead of the parametric model used to train the neural network.

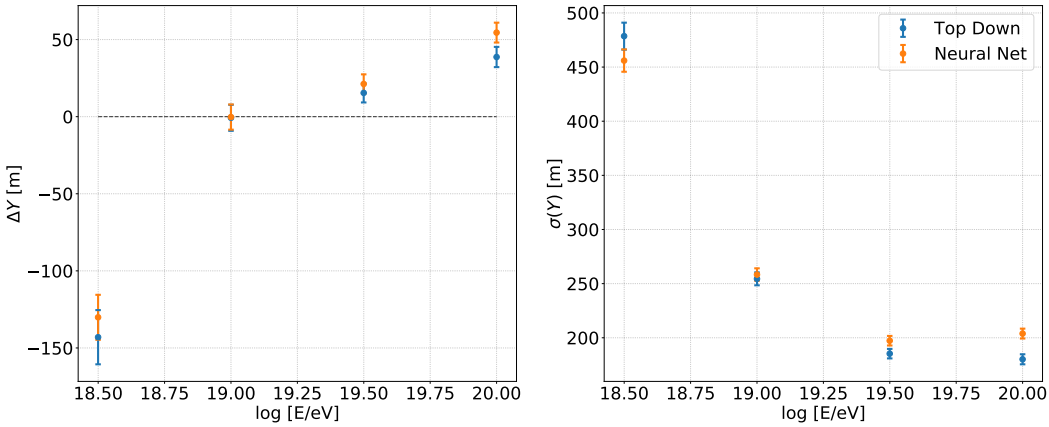


Figure 9.40. The bias (left) and resolution (right) of the y component of the core position as a function of energy for the top-down reconstruction (blue) and the neural network reconstruction (orange). These events were simulated with the January molecular model instead of the parametric model used to train the neural network.

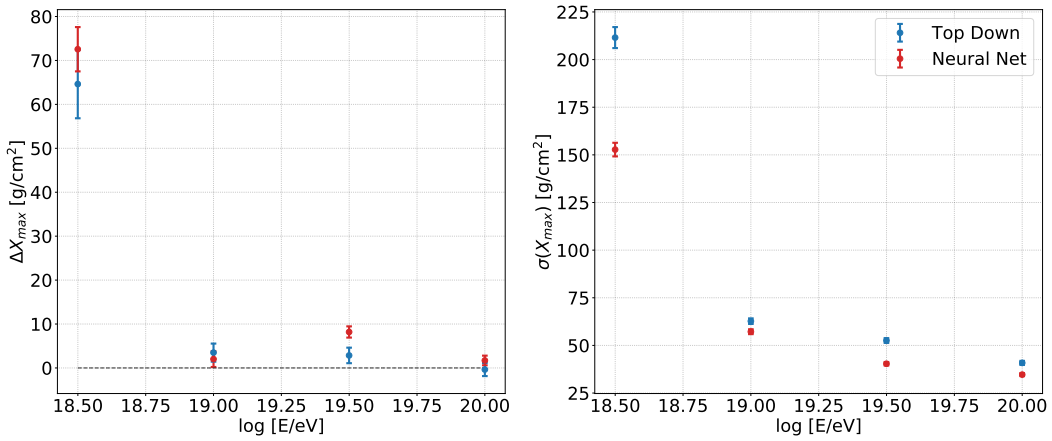


Figure 9.41. The $X_{\max}$ bias (left) and resolution (right) as a function of energy for the top-down reconstruction (blue) and the neural network reconstruction (red). These events were simulated with an increased VAOD compared to the simulated events with which the neural network was trained.

with an increase in the VAOD from ∼0.055 to 0.1. The $X_{\max}$ reconstruction bias was relatively unaffected by the change in the VAOD, as discussed in Section 8.7. This is likely due to the fact that most of the aerosol content of the atmosphere is relatively close to the ground. Therefore, any changes in the shower profile do not significantly change the position of $X_{\max}$ relative to the detector. The $X_{\max}$ reconstruction bias of the first guess and the subsequent top-down reconstruction is shown in Figure 9.41. The bias is relatively small, especially above 10^{19} eV. The energy bias, on the other hand, depends significantly on the aerosol content of the atmosphere since it can significantly change the transmission of fluorescence light from the shower to the detector. Shown in Figure 9.42 is a comparison of the biased first guess of energy from the neural network to the subsequent top-down reconstruction. As for the changes to the molecular model, the top-down reconstruction is able to recover an accurate estimate of the energy, even with a biased first guess. Again. this further validates the top-down reconstruction.

As for the changes in the molecular atmosphere, the zenith angle and the core position are relatively unaffected by changes in the aerosol model. The reconstruction bias and resolution of the zenith angle and the x and y components of the core position are included in Figures 9.43 to 9.45. The differences in the reconstruction bias are insignificant both for the top-down reconstruction, and the neural network reconstruction as discussed in Section 8.7.

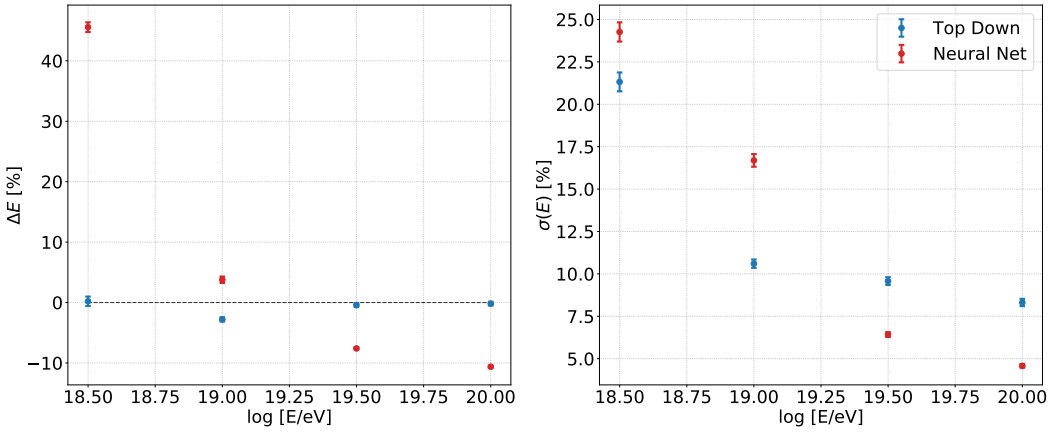


Figure 9.42. The relative energy bias (left) and resolution (right) as a function of energy for the top-down reconstruction (blue) and the neural network reconstruction (red). These events were simulated with an increased VAOD compared to the simulated events with which the neural network was trained.

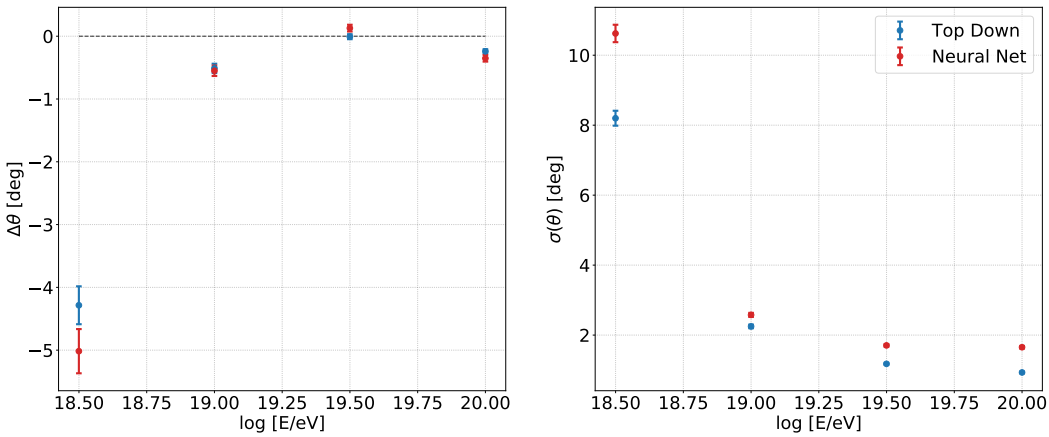


Figure 9.43. The zenith angle bias (left) and resolution (right) as a function of energy for the top-down reconstruction (blue) and the neural network reconstruction (red). These events were simulated with an increased VAOD compared to the simulated events with which the neural network was trained.

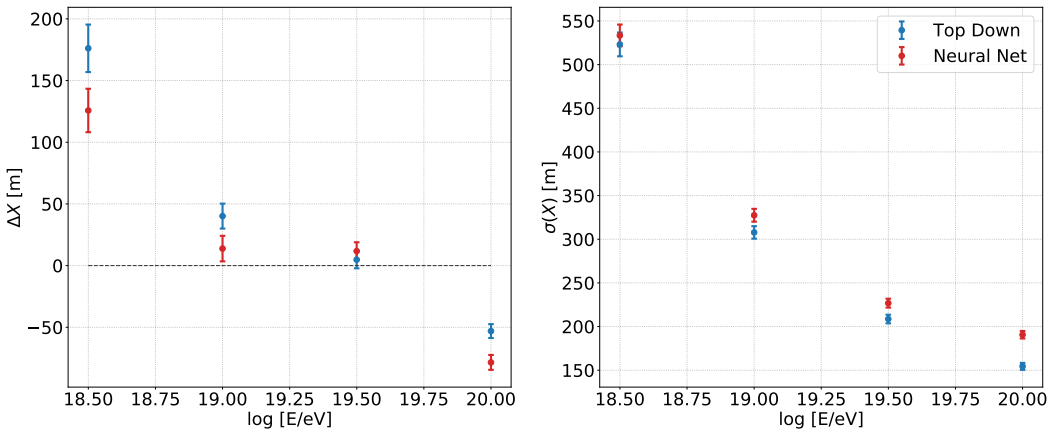


Figure 9.44. The bias (left) and resolution (right) of the x component of the core position as a function of energy for the top-down reconstruction (blue) and the neural network reconstruction (red). These events were simulated with an increased VAOD compared to the simulated events with which the neural network was trained.

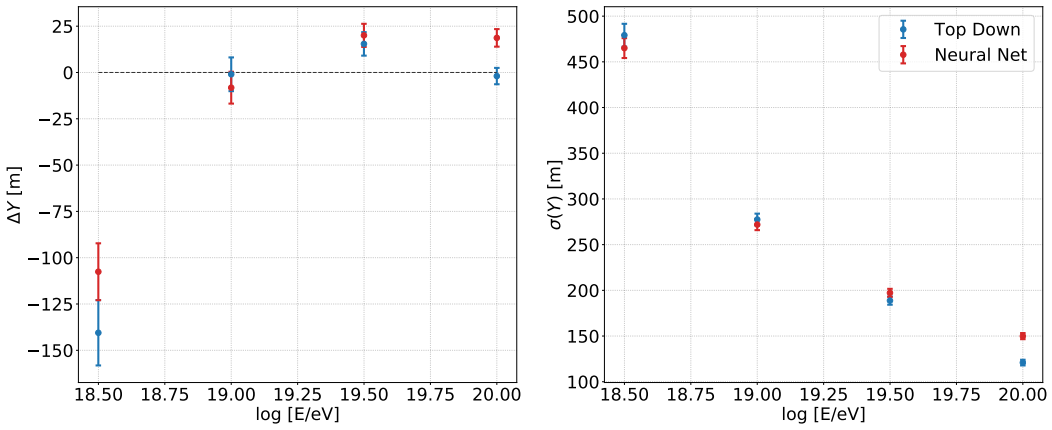


Figure 9.45. The bias (left) and resolution (right) of the y component of the core position as a function of energy for the top-down reconstruction (blue) and the neural network reconstruction (red). These events were simulated with an increased VAOD compared to the simulated events with which the neural network was trained.

9.3 Reconstruction of Real Events

This section considers the reconstruction of real events measured by FAST prototype telescopes in coincidence with TA. Whilst it is not possible to study the array reconstruction mode, since the FAST prototype telescopes have yet to measure air showers in stereo mode, these case studies provide some insight to how the developed reconstruction algorithms perform on measured data with the hybrid reconstruction mode. To this end, two case studies of event reconstructions will be considered with the geometry fixed to the result from the TA monocular reconstruction. The reconstructed shower parameters can be compared to those of the TA monocular FD reconstruction or the TA SD reconstruction, depending on their availabilities.

Event 1: 15$^{\text{th}}$ of May 2018

This event was first discussed in Section 6.2.3. It was measured on the 15$^{\text{th}}$ of May 2018 in-coincidence with the TA FD. According to the TA monocular reconstruction, the event has an energy of $\sim$19 EeV, an X_{max} value of 852 g/cm^2 and a zenith angle of $\sim$55°. The impact parameter for this event is 6 km. In order to reconstruct this event with the top-down reconstruction, the geometry is fixed to that obtained from TA. The simulated traces from the best-fit shower parameters are compared to the measured traces in Figure 9.46. There is generally good agreement between the shape of the measured and simulated traces, with the exception of one of the lower PMTs. This difference can likely be explained by uncertainties in the FAST calibration or uncertainties in the reconstructed geometry obtained by FAST. Additionally, the orientation of the shower axis for this event is such that the measured signals have a large contribution from direct Cherenkov light. This can lead to large systematic uncertainties in the shower geometry. The reconstructed energy and X_{max} are 17$\pm$1.5 EeV and 843$\pm$55 g/cm^2, respectively. A systematic energy bias of $\sim$14% between FAST and TA is expected due to the different fluorescence yield models used by the two experiments [28, 217]. Therefore, the reconstructed X_{max} and energy from the top-down reconstruction are compatible with those obtained from the TA reconstruction within systematic uncertainty.

Event 2: 10$^{\text{th}}$ of January 2019

This event was measured by two of the FAST prototype telescopes in-coincidence with TA on the 10$^{\text{th}}$ of January 2019. According to the TA monocular reconstruction, the event also has an energy of $\sim$19 EeV, an X_{max} value of 824 g/cm^2 and a zenith angle of $\sim$33°. Although this event was reconstructed at approximately the same energy as the first event, the impact parameter is more than double at a distance of 13 km. Therefore a significantly smaller SNR is expected in the measured FAST pulses. The simulated traces from the best-fit shower parameters are compared to

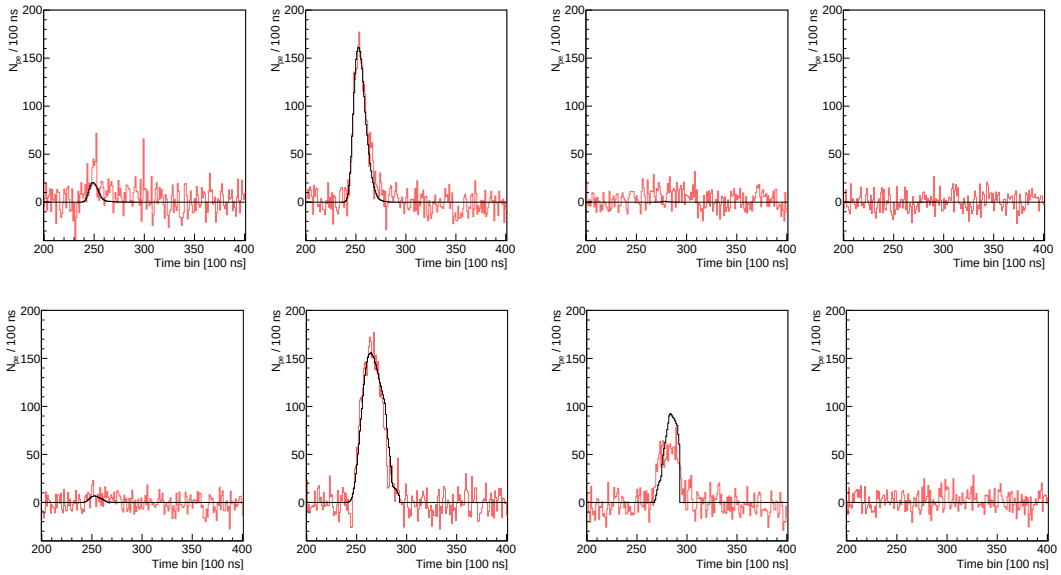


Figure 9.46. A comparison between the simulated traces from the best-fit shower parameters (black) and the measured FAST traces (red). The PMT traces are arranged in a sky-view orientation. This event was measured by the FAST prototype telescopes at TA on the 15$^{\text{th}}$ of May 2018.

the measured traces in Figure 9.47. There is generally good agreement between the shape of the measured and simulated traces. Further understanding of the FAST telescope calibration and optics are required to reduce the discrepancy and ensure the detector simulation is as realistic as possible. An energy of 19$\pm$3 EeV and an X_{max} of 808$\pm$70 g/cm^2 are obtained from the top-down reconstruction with the shower geometry fixed to the result of the TA reconstruction. The reconstructed X_{max} and energy from the top-down reconstruction are compatible with those obtained from the TA reconstruction within statistical uncertainty. The success of these example events gives further credence to the top-down reconstruction as a viable option for the FAST experiment.

9.4 Conclusions

The performance of the top-down reconstruction has been evaluated in two independent modes. The first being the hybrid reconstruction mode where only X_{max} and energy are determined by the FAST telescopes, using a fixed SD geometry obtained from an SD. The second mode is the array reconstruction mode where FAST stations are arranged in a triangular configuration with a spacing of 20 km. Overall, the top-down reconstruction performs especially well above 10^{19} eV, with significantly reduced performance at lower energy.

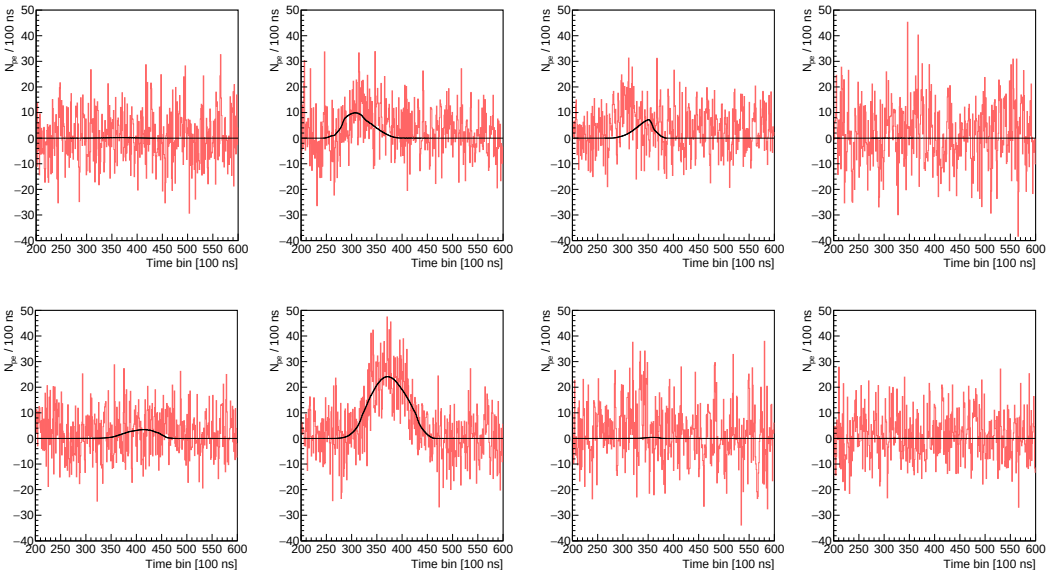


Figure 9.47. A comparison between the simulated traces from the best-fit shower parameters (black) and the measured FAST traces (red). The PMT traces are arranged in a sky-view orientation. This event was measured by the FAST prototype telescopes at TA on the 10$^{\text{th}}$ of January 2019.

The hybrid reconstruction mode has been shown to be successful for simulated events. The resolution of the hybrid reconstruction is best at core distances of 6-8 km where an X_{max} resolution of $\sim$50 g/cm^2 can be obtained above 10^{19} eV. The resolution significantly degrades at smaller core distances for a given energy, due to the limited FoV of the FAST telescopes. This could potentially be improved by adding additional FAST telescopes with a higher elevation FoV similar to the High-elevation Auger telescopes (HEAT) at the Pierre Auger Observatory (Auger). Future simulation studies of the top-down reconstruction could investigate this possibility. The resolution is also degraded at large core distances due to lower SNR.

The top-down reconstruction in the full array mode has also shown to be viable. However, there are significant issues with reduced reconstruction efficiency at high energy due to imprecision of the detector simulation. Future improvements of the simulation or the minimiser should be considered to provide the best reconstruction performance. A scaled version of the likelihood function was used to improve the reconstruction efficiency at the expense of resolution in the reconstructed shower parameters. The top-down reconstruction with a scaled likelihood function can achieve an X_{max} resolution of 50-60 g/cm^2, an energy resolution of 12-15%, an angular resolution of 2-3°, and a core resolution of 200-300 m at the FAST target energy of $10^{19.5}$ eV. With improvements to the precision of the detector simulation, the top-down reconstruction resolution can be significantly improved; the top-down reconstruction with the unscaled likelihood function can achieve an X_{max} resolution of $\sim$30 g/cm^2, an energy resolution of 6-8%, a zenith angle resolution of $\sim$0.7°, and

a core resolution of $\sim$150 m at the FAST target energy of $10^{19.5}$ eV. The top-down reconstruction has also been shown to successfully recover from a biased first guess of the shower parameters from the neural network reconstruction, further validating the entire FAST reconstruction chain.

Two real events measured by the FAST prototype telescopes at TA have been successfully reconstructed using the hybrid reconstruction mode with a shower geometry fixed to that obtained by the TA reconstruction. Further investigation of the discrepancies between the detector simulation and measurements should be considered. It is essential for the detector simulation to be as realistic as possible for the top-down reconstruction. Future FAST prototype telescopes will be arranged in a stereo configuration. It will be important to test the top-down reconstruction in this configuration to determine the performance of the geometry reconstruction.

Future studies should also consider a more sophisticated trigger and event selection algorithm. It will be critical to identify events which may trigger only one or two FAST stations, but a third may measure sub-threshold signals. This information may still be useful for the top-down reconstruction. Efficient communication between FAST stations will be crucial for the successful operation of a future FAST array.

10 Conclusions

This thesis has focused on extending the energy range of ultra-high energy cosmic ray (UHECR) measurements, with a particular focus on fluorescence detectors. This was achieved in two ways. Firstly, by removing the fiducial volume cuts and instead correcting for the X_{max} acceptance in order to sample unbiased X_{max} distributions. This procedure has allowed the extension of the Pierre Auger Observatory (Auger) mass composition analysis to both lower and higher energies. Secondly, by developing a novel technique for the reconstruction of extensive air showers (EAS) using the Fluorescence detector Array of Single-pixel Telescopes (FAST).

Mass Composition Studies at the Pierre Auger Observatory

The energy threshold of the HEAT-Coihueco (HeCo) X_{max} analysis can be lowered using the new Surface Detector (SD) triggers, which are more sensitive to low energy showers in the 750 m array. Parametrisations of the trigger probability of the SD were derived in Chapter 4. The trigger probability was used in conjunction with a series of modified selection cuts to increase the number of events in the data set. A modified analysis technique using a kernel function to fit the X_{max} detector acceptance from Monte Carlo (MC) simulated events was developed in Chapter 5. The estimated X_{max} detector acceptance replaced the fiducial volume cuts, which are responsible for significantly reducing statistics of the standard analysis. The systematic uncertainties of the new technique were studied in detail. Very good agreement was shown between the results of the standard X_{max} analysis and the results of this work. However systematic differences (below $\sim 10^{17.5}$ eV) between fluorescence measurements and measurements from other experiments warrants further investigation. Additional cross-checks of the X_{max} analysis techniques should be considered. With future additional data at the highest energies (which may be obtained with a FAST array), significant astrophysical implications may be uncovered with the knowledge of the mass composition of UHECRs.

The updated X_{max} moments in Figure 5.42 represent significant extensions to the Auger mass composition analysis energy range. The new analysis presented in this thesis has provided a four additional energy bins between $10^{16.8}$ and $10^{17.2}$ eV, as well as one additional energy bin at the highest energies (above $10^{19.8}$ eV). The agreement with the standard analysis in the overlapping range is extremely good over the full energy range from $10^{17.2}$ to $\sim$ $10^{19.7}$ eV. Possible systematic

differences between fluorescence and Cherenkov detectors have been highlighted by the extension to lower energies, while the extension to higher energy provides the possibility of uncovering significant astrophysical implications with additional data in the future.

The Fluorescence detector Array of Single-pixel Telescopes (FAST)

FAST is a prototype design for a next generation cosmic ray observatory which will be capable of measuring the highest energy cosmic rays with an unprecedented exposure. A proof-of-concept reconstruction technique based on a realistic detector simulation has been developed in this thesis, including a full event data structure, which is now being used by the FAST collaboration.

The top-down reconstruction is capable of reconstructing EAS in two distinct modes; the "hybrid" mode and the array mode. In the hybrid mode FAST telescopes are employed to reconstruct X_{max} and energy given an independent estimate of the shower geometry from a companion SD. In the hybrid mode the top-down reconstruction is capable in its own right. In the array mode, however, the top-down reconstruction must be provided with a first guess of the shower parameters in order to converge.

A first guess reconstruction was designed using a neural network trained on characteristics of the individual pulses measured in the cameras of the FAST telescopes. The neural network reconstruction performs extremely well in the ideal case; however, it is highly susceptible to systematic biases which arise due to the fixed detector state (including atmosphere) in the simulated data set used for training. However, the top-down reconstruction is capable of recovering these systematic uncertainties since changes in the detector state can be accounted for in the detector simulation based on independent measurements. However, the top-down reconstruction relies heavily on an accurate and realistic detector simulation. The reconstruction efficiency is limited at very high energy due to inaccuracies in the detector simulation which prevent the minimisation of the likelihood function. Future work should consider improvements to the detector simulation or the minimisation algorithm to circumvent these issues.

Following successful tests with the prototype FAST telescopes at the Telescope Array (TA) and Auger, future FAST prototype telescopes should be configured to test the stereoscopic operation of FAST in the array mode. Such tests will give further insight to the performance of the reconstruction algorithms on real measured events. An important challenge for FAST will be the development of a trigger algorithm which will be capable of detecting air showers in the array mode.

List of Acronyms

ADC analogue-to-digital converter

AERA Auger Engineering Radio Array

AGASA the Akeno Giant Air-Shower Array

AGN active galactic nuclei

AMIGA Auger Muons and Infill for the Ground Array

Auger the Pierre Auger Observatory

BRM Black Rock Mesa

CDAS the Central Data Acquisition System

CIC Constant Intensity Cut

CLF Central Laser Facility

CMB cosmic microwave background

DAQ data acquisition

DSA diffusive shock acceleration

EAS extensive air shower

EMG exponentially modified Gaussian

FASCam FAST all sky camera

FAST the Fluorescence detector Array of Single-pixel Telescopes

FD Fluorescence Detector

FLT First Level Trigger

FoV field of view

GDAS the Global Data Assimilation System

GOES the Geostationary Operational Environmental Satellite system

GRB gamma-ray burst

GZK Greisen-Zatsepin-Kuz'min

HEAT the High-elevation Auger telescopes

HeCo HEAT-Coihueco

HiRes High Resolution Fly's Eye

HV high voltage

IGM intergalactic medium

ISM interstellar medium

LDF lateral distribution function

LHC Large Hadron Collider

LOFAR the Low Frequency Array

LR Long Ridge

LTP Lateral Trigger Probability

LV low voltage

MC Monte Carlo

MD Middle Drum

MoPS multiplicity of positive steps

NKG Nishimura-Kamata-Greisen

NSB night-sky background

PDF probability density function

PMT photomultiplier tube

PSF point spread function

SD Surface Detector

SDP shower-detector plane

SLT Second Level Trigger

SNR signal-to-noise ratio

SQM sky quality monitor

SSD Surface Scintillation Detector

SUGAR the Sydney University Giant Air-shower Recorder

TA the Telescope Array

TALE Telescope Array Low-energy Extension

Th threshold

TLT Third Level Trigger

ToT time-over-threshold

ToTd time-over-threshold de-convoluted

UHECR ultra-high energy cosmic ray

UMD Underground Muon Detector

UV ultra-violet

VAOD vertical aerosol optical depth

VEM vertical equivalent muon

WCD Water Cherenkov Detector

XLF eXtreme Laser Facility

YAP Yttrium-Aluminium-Perovskite

A

LTP Parametrisations

A.1 Parametrisation Coefficients

The Lateral Trigger Probability (LTP) is derived, as discussed in Section 4.2 using the following fit function:

$$LTP(r) = \begin{cases} \frac{1}{1+e^{\frac{r-R_0}{\Delta R}}} & \text{for } r \leq R_0 \\ \frac{1}{2e^{\frac{r-R_0}{2\Delta R}}} & \text{for } r > R_0 \end{cases} \tag{A.1}$$

where R_0 is the station-axis distance where the LTP is equal to 0.5.

The dependencies of the fit parameters on energy and zenith angle are discussed in Section 4.2.4. The dependencies are parametrised using a 2-dimensional polynomial in $\cos\theta$ and $\log_{10} E$. The best fit coefficients for each of the trigger configurations are summarised below in a series of matrix equations. An electronic version has been implemented in the Offline software framework and can be provided on request.

ToT

$$\frac{R_0}{\text{km}} = \begin{pmatrix} 1 \\ \cos\theta \\ \cos^2\theta \end{pmatrix}^T \cdot \left[\begin{pmatrix} 54.6674 & -7.38619 & 0.239342 \\ -75.9798 & 10.0703 & -0.312338 \\ 45.8044 & -5.87679 & 0.178285 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \log E \\ \log^2 E \end{pmatrix} \right]$$

$$\frac{\Delta R}{\text{km}} = \begin{pmatrix} 1 \\ \cos\theta \\ \cos^2\theta \end{pmatrix}^T \cdot \left[\begin{pmatrix} -9.25339 & 0.846544 & -0.0175944 \\ 19.6017 & -1.75393 & 0.0356426 \\ -9.12718 & 0.750772 & -0.0128803 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \log E \\ \log^2 E \end{pmatrix} \right]$$

ToT || TH-T1 || TH-T2

$$\frac{R_0}{\mathrm{km}} = \begin{pmatrix} 1 \\ \cos\theta \\ \cos^2\theta \end{pmatrix}^T \cdot \left[\begin{pmatrix} 54.1579 & -7.34736 & 0.241278 \\ -75.2313 & 10.0333 & -0.318171 \\ 45.6364 & -5.89077 & 0.182705 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \log E \\ \log^2 E \end{pmatrix} \right]$$

$$\frac{\Delta R}{\mathrm{km}} = \begin{pmatrix} 1 \\ \cos\theta \\ \cos^2\theta \end{pmatrix}^T \cdot \left[\begin{pmatrix} -8.15407 & 0.839128 & -0.0200479 \\ 17.3503 & -1.7675 & 0.042336 \\ -7.8054 & 0.754822 & -0.0166769 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \log E \\ \log^2 E \end{pmatrix} \right]$$

ToT || TH-T1 || TH-T2 || ToTd || MoPS

$$\frac{R_0}{\mathrm{km}} = \begin{pmatrix} 1 \\ \cos\theta \\ \cos^2\theta \end{pmatrix}^T \cdot \left[\begin{pmatrix} 84.0378 & -10.3988 & 0.316334 \\ -84.7232 & 10.203 & -0.292112 \\ 19.0006 & -2.39227 & 0.0675405 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \log E \\ \log^2 E \end{pmatrix} \right]$$

$$\frac{\Delta R}{\mathrm{km}} = \begin{pmatrix} 1 \\ \cos\theta \\ \cos^2\theta \end{pmatrix}^T \cdot \left[\begin{pmatrix} -8.40531 & 0.813001 & -0.0178409 \\ 18.8802 & -1.78354 & 0.0385972 \\ -8.51627 & 0.729046 & -0.0130695 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \log E \\ \log^2 E \end{pmatrix} \right]$$

ToT || ToTd || MoPS

$$\frac{R_0}{\mathrm{km}} = \begin{pmatrix} 1 \\ \cos\theta \\ \cos^2\theta \end{pmatrix}^T \cdot \left[\begin{pmatrix} 96.9204 & -11.6599 & 0.344498 \\ -90.1559 & 10.3637 & -0.279465 \\ 8.76524 & -0.96961 & 0.017364 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \log E \\ \log^2 E \end{pmatrix} \right]$$

$$\frac{\Delta R}{\mathrm{km}} = \begin{pmatrix} 1 \\ \cos\theta \\ \cos^2\theta \end{pmatrix}^T \cdot \left[\begin{pmatrix} 0.00212581 & 0.031512 & -0.000889402 \\ 14.2694 & -1.69864 & 0.0491729 \\ -14.2941 & 1.65644 & -0.0473367 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \log E \\ \log^2 E \end{pmatrix} \right]$$

ToT || TH-T2

$$\frac{R_0}{\mathrm{km}} = \begin{pmatrix} 1 \\ \cos\theta \\ \cos^2\theta \end{pmatrix}^T \cdot \left[\begin{pmatrix} 54.5695 & -7.39799 & 0.240417 \\ -75.4021 & 10.0638 & -0.314029 \\ 45.4676 & -5.87553 & 0.179414 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \log E \\ \log^2 E \end{pmatrix} \right]$$

$$\frac{\Delta R}{\mathrm{km}} = \begin{pmatrix} 1 \\ \cos\theta \\ \cos^2\theta \end{pmatrix}^T \cdot \left[\begin{pmatrix} -9.40454 & 0.851313 & -0.0173854 \\ 19.8403 & -1.75004 & 0.0346752 \\ -9.31427 & 0.752584 & -0.0123946 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \log E \\ \log^2 E \end{pmatrix} \right]$$

ToT || TH-T2 || ToTd || MoPS

$$\frac{R_0}{\mathrm{km}} = \begin{pmatrix} 1 \\ \cos\theta \\ \cos^2\theta \end{pmatrix}^T \cdot \left[\begin{pmatrix} 96.9455 & -11.6824 & 0.345714 \\ -89.8089 & 10.3729 & -0.281165 \\ 8.37675 & -0.955169 & 0.0178299 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \log E \\ \log^2 E \end{pmatrix} \right]$$

$$\frac{\Delta R}{\mathrm{km}} = \begin{pmatrix} 1 \\ \cos\theta \\ \cos^2\theta \end{pmatrix}^T \cdot \left[\begin{pmatrix} -3.17923 & 0.310137 & -0.00643927 \\ 15.9169 & -1.68303 & 0.0429473 \\ -12.4797 & 1.32801 & -0.0345235 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \log E \\ \log^2 E \end{pmatrix} \right]$$

A.2 LTP Fits

The measured LTP data and fits using Equation (4.7) for each of the trigger conditions are plotted below in various energy and zenith angle bins. The fit generally does well with the exception of the large zenith angle and low energy combination, where the data is more difficult to fit. This is due to a lack of statistics (apparent from the larger vertical error bars) for this combination of zenith angle and energy. A lack of statistics is also observed for the first bin of station-axis distance in most cases. A number of these stations do trigger in the data, however the hybrid station (largest signal) is always ignored in the LTP analysis. This is discussed in Section 4.2.

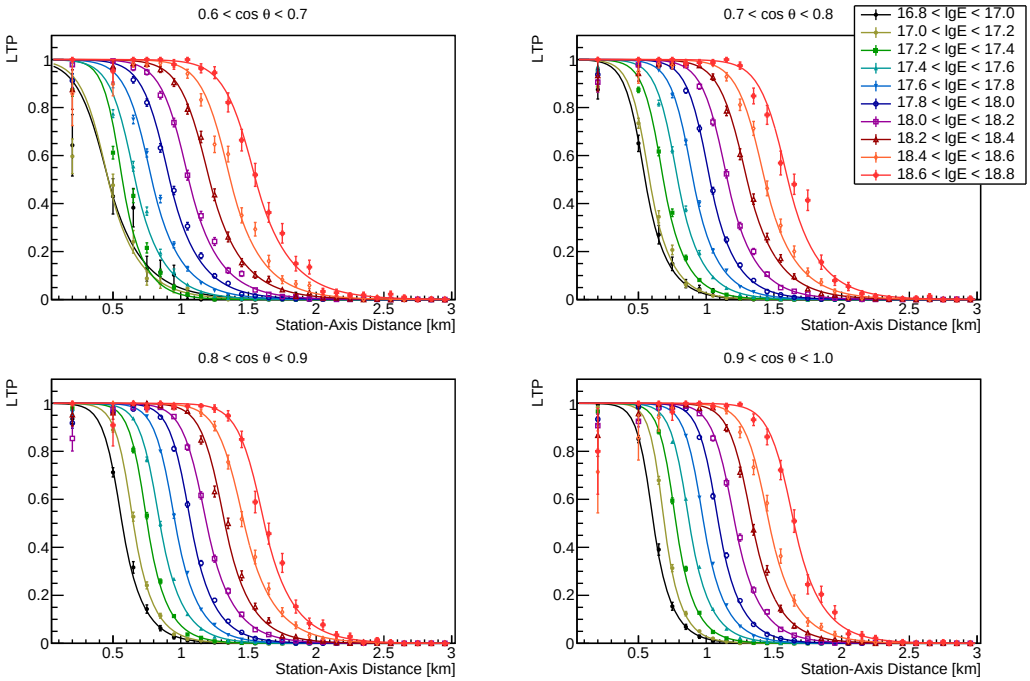


Figure A.1. The measured LTP data and fits for ToT triggered stations.

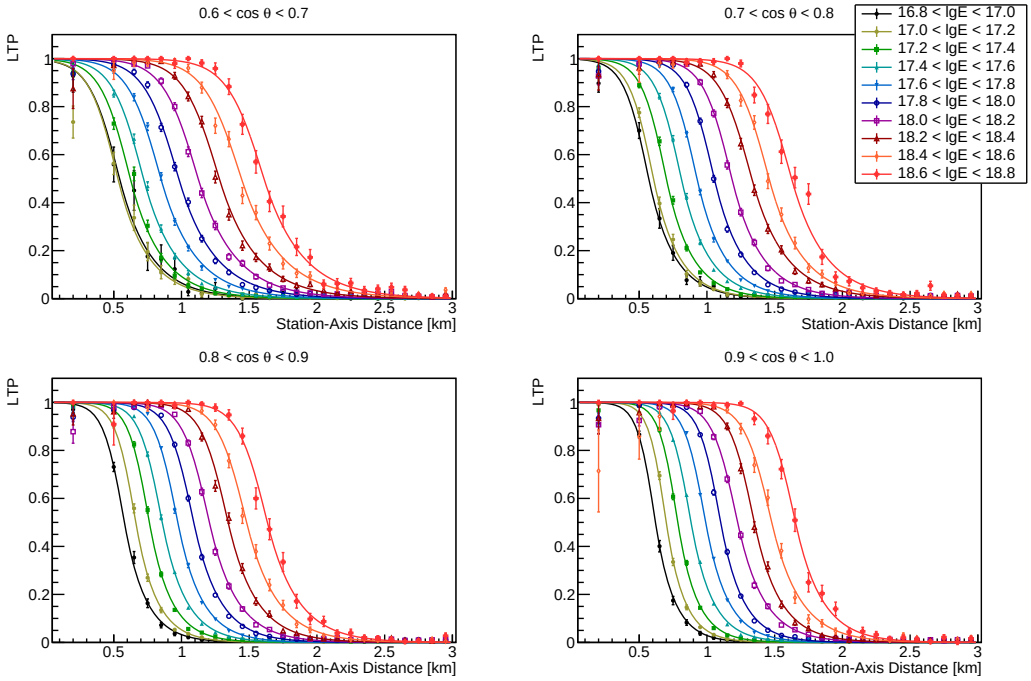


Figure A.2. The measured LTP data and fits for ToT || TH-T1 || TH-T2 triggered stations.

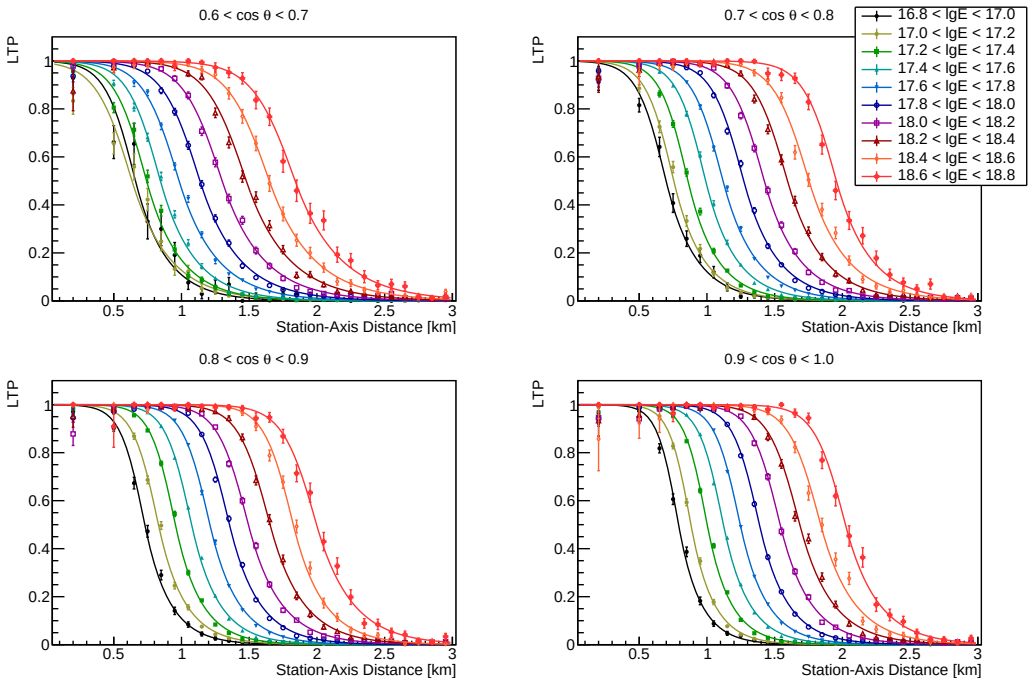


Figure A.3. The measured LTP data and fits for ToT || TH-T1 || TH-T2 || ToTd || MoPS triggered stations.

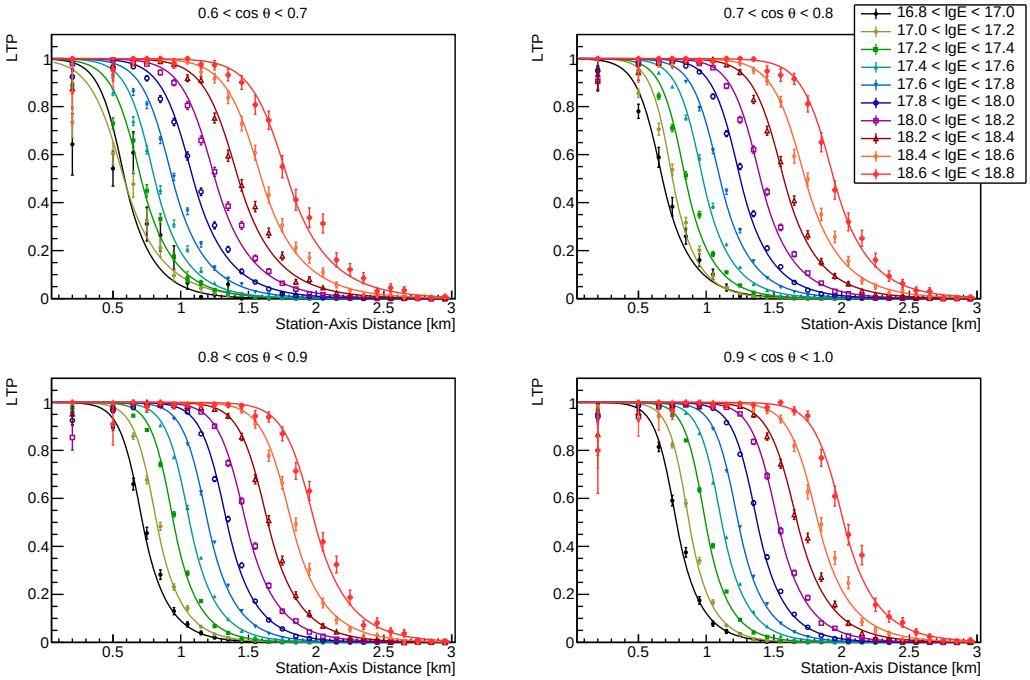


Figure A.4. The measured LTP data and fits for ToT || ToTd || MoPS triggered stations.

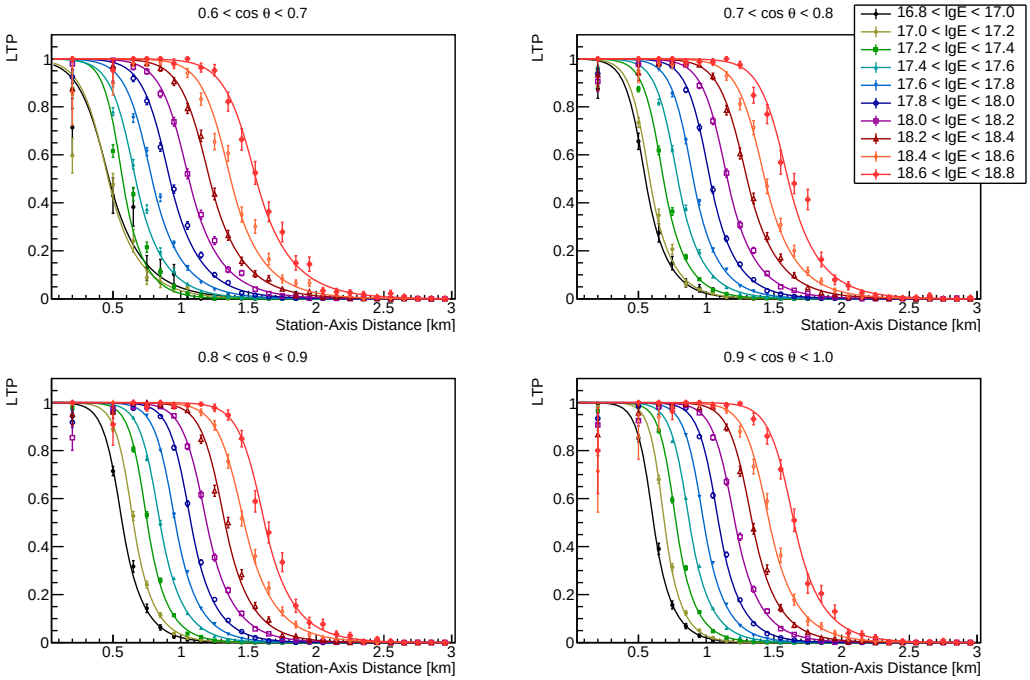


Figure A.5. The measured LTP data and fits for ToT || TH-T1 || TH-T2 triggered stations.

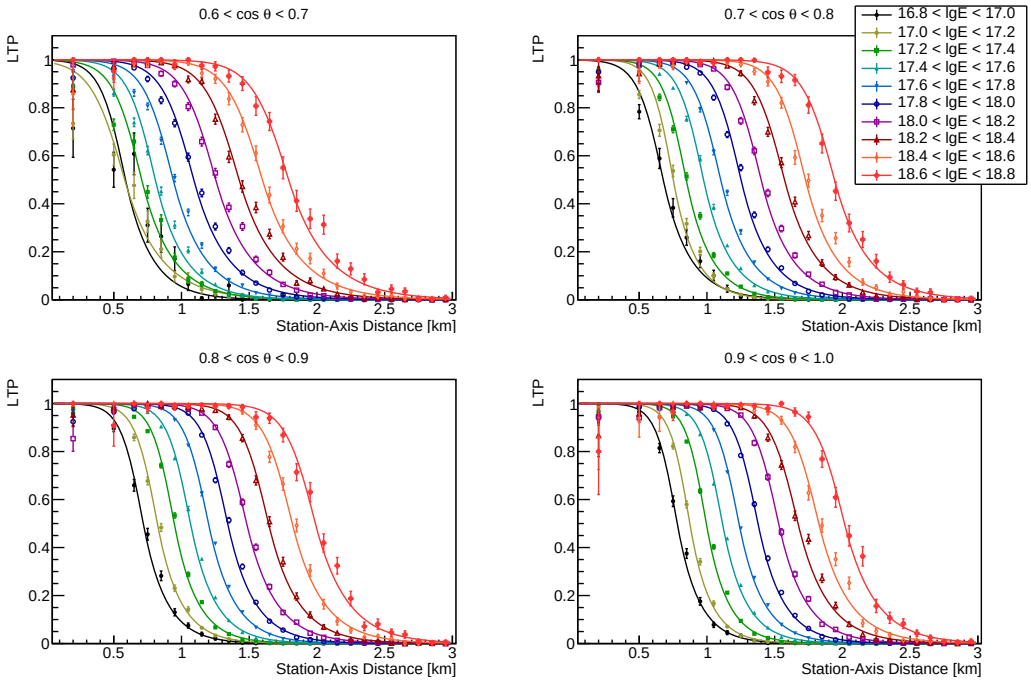


Figure A.6. The measured LTP data and fits for ToT || TH-T2 || ToTd || MoPS triggered stations.

A.3 Dependence of Fit Parameters on Energy and Zenith Angle

Parameters R_0 and ΔR are plotted below as a function of $\cos\theta$ and $\log E$ for each of the trigger conditions. The red surface shows the result of the quadratic fit described in Section 4.2.4. The coefficients of these fits are shown in Appendix A.1. Note the scale of R_0 varies by hundreds of meters while the scale of ΔR varies by only tens of meters in the fitted range. This suggests R_0 dominates the energy and zenith angle dependence of the LTP.

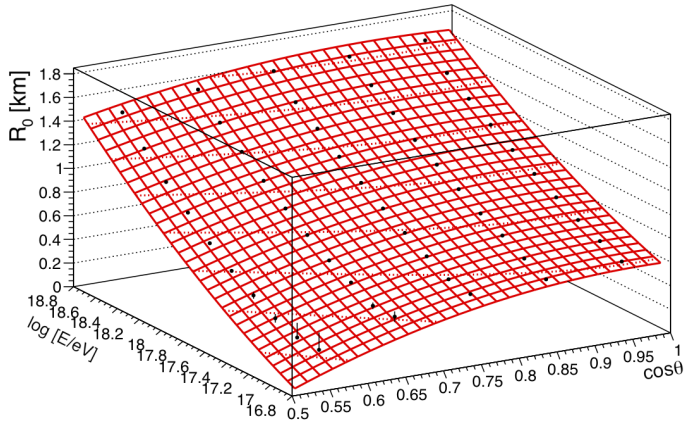


Figure A.7. The fitted R_0 data as a function of $\cos\theta$ and $\log E$ along with the quadratic fits for ToT triggered stations.

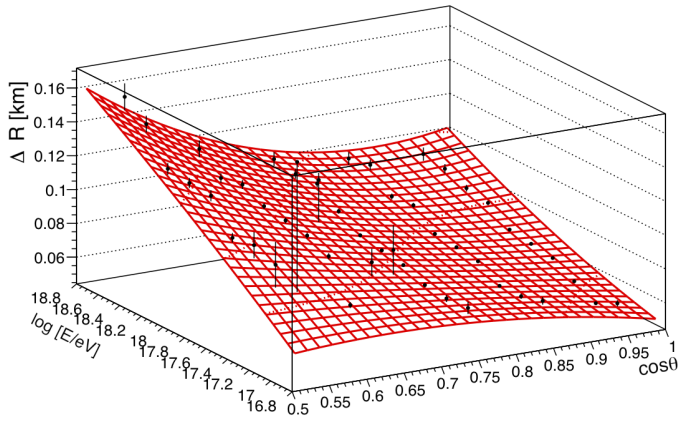


Figure A.8. The fitted ΔR data as a function of $\cos\theta$ and $\log E$ along with the quadratic fits for ToT triggered stations.

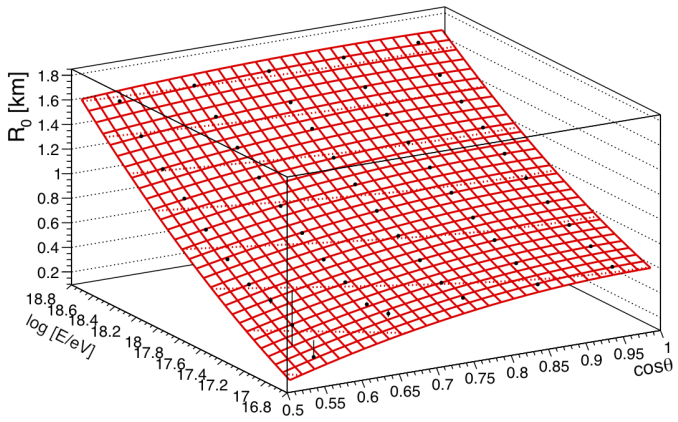


Figure A.9. The fitted R_0 data as a function of $\cos\theta$ and $\log E$ along with the quadratic fits for ToT || TH-T1 || TH-T2 triggered stations.

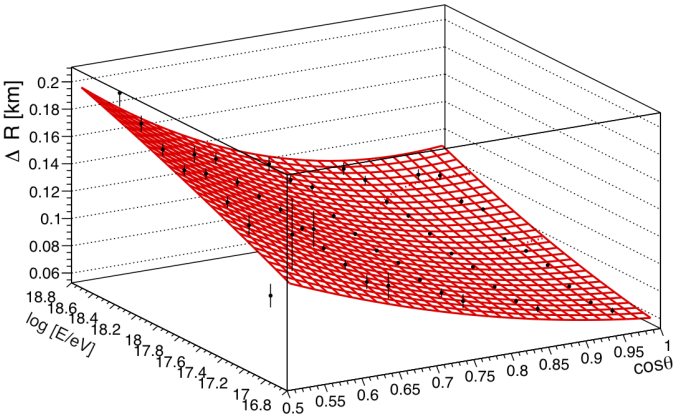


Figure A.10. The fitted ΔR data as a function of $\cos\theta$ and $\log E$ along with the quadratic fits for ToT || TH-T1 || TH-T2 triggered stations.

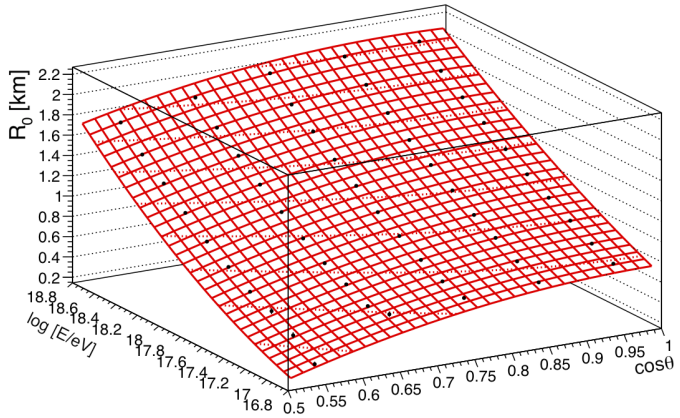


Figure A.11. The fitted R_0 data as a function of $\cos\theta$ and $\log E$ along with the quadratic fits for ToT || TH-T1 || TH-T2 || ToTd || MoPS triggered stations.

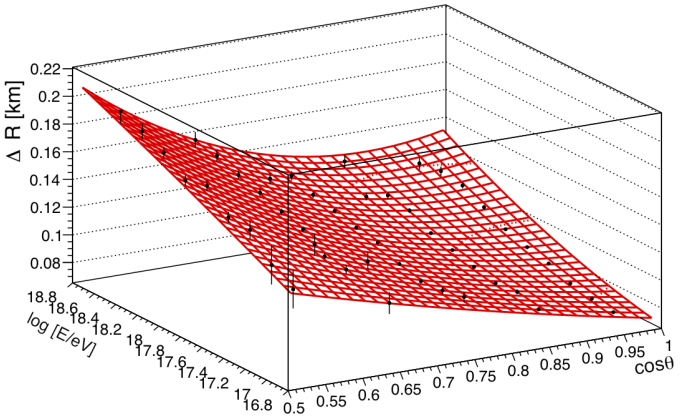


Figure A.12. The fitted ΔR data as a function of $\cos\theta$ and $\log E$ along with the quadratic fits for ToT || TH-T1 || TH-T2 || ToTd || MoPS triggered stations.

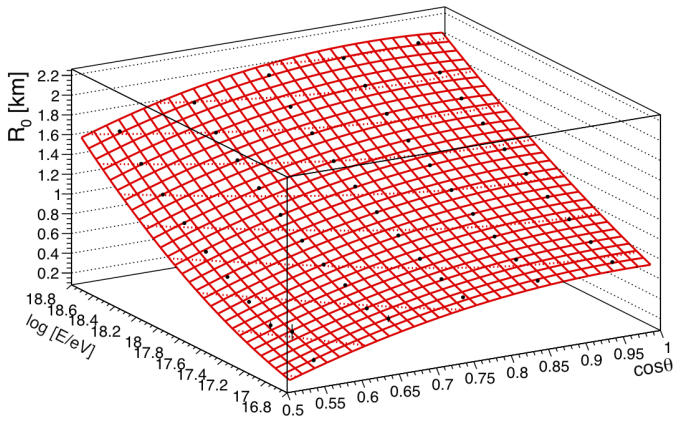


Figure A.13. The fitted R_0 data as a function of $\cos\theta$ and $\log E$ along with the quadratic fits for ToT || ToTd || MoPS triggered stations.

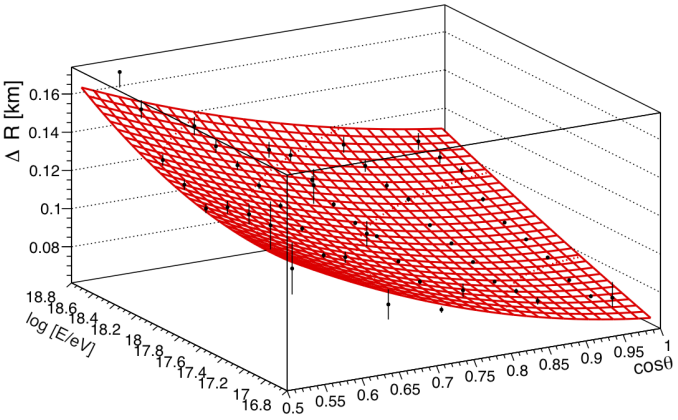


Figure A.14. The fitted ΔR data as a function of $\cos\theta$ and $\log E$ along with the quadratic fits for ToT || ToTd || MoPS triggered stations.

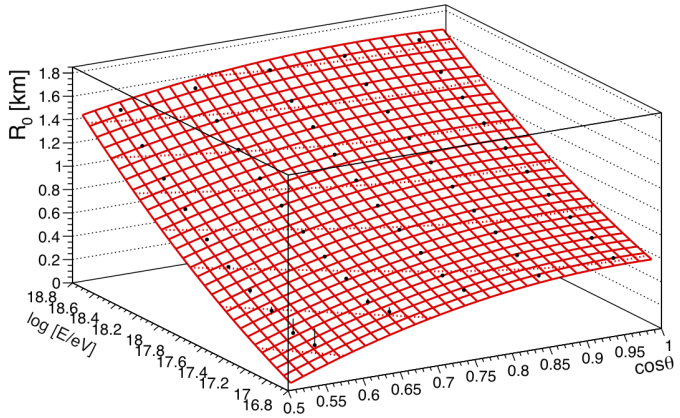


Figure A.15. The fitted R_0 data as a function of $\cos\theta$ and $\log E$ along with the quadratic fits for ToT || TH-T1 || TH-T2 triggered stations.

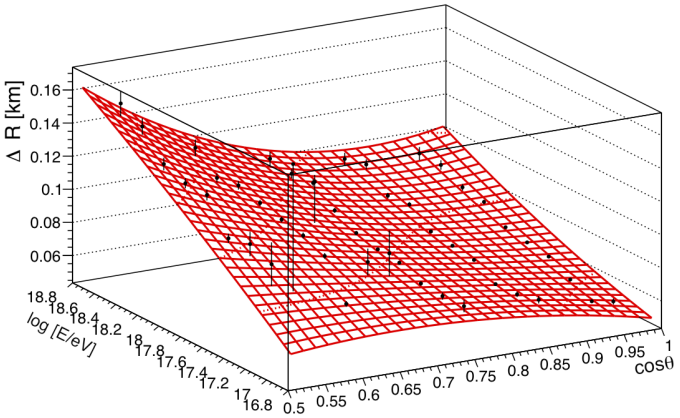


Figure A.16. The fitted ΔR data as a function of $\cos\theta$ and $\log E$ quadratic fits for ToT || TH-T1 || TH-T2 triggered stations.

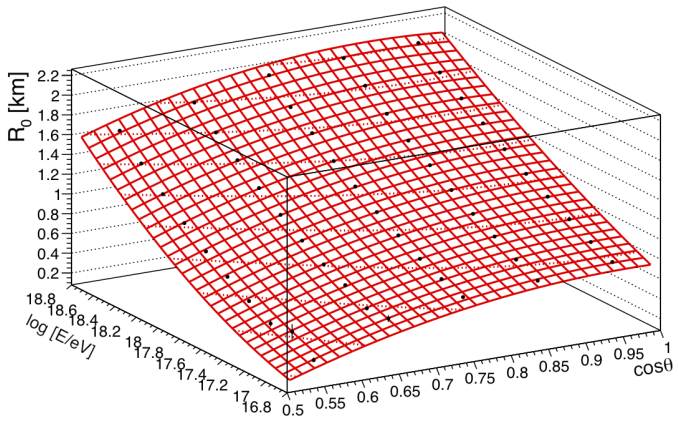


Figure A.17. The fitted R_0 data as a function of $\cos\theta$ and $\log E$ quadratic fits for ToT || TH-T2 || ToTd || MoPS triggered stations.

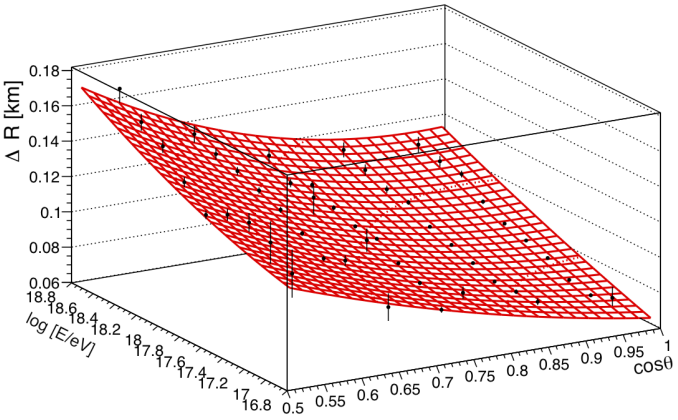


Figure A.18. The fitted ΔR data as a function of $\cos\theta$ and $\log E$ quadratic fits for ToT || TH-T2 || ToTd || MoPS triggered stations.

A.4 Comparison of Parametrisation to Data

The measured LTP data are compared with the final parametrisations for each of the trigger conditions below, in various energy and zenith angle bins. The data agrees with the parametrisation in most cases, with the exception of low energy, large zenith angle showers where statistics are poor. The parametrisations, however, are conservative. That is, the parametrisation predicts an LTP that is less than the data.

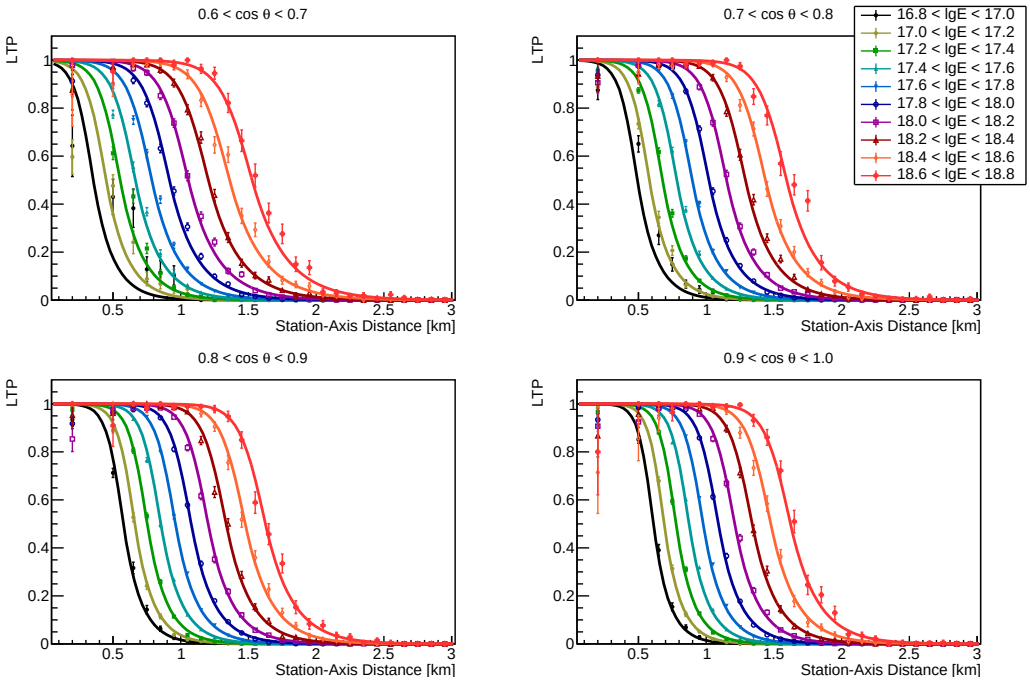


Figure A.19. The measured LTP data compared with the final parametrisation for ToT triggered stations.

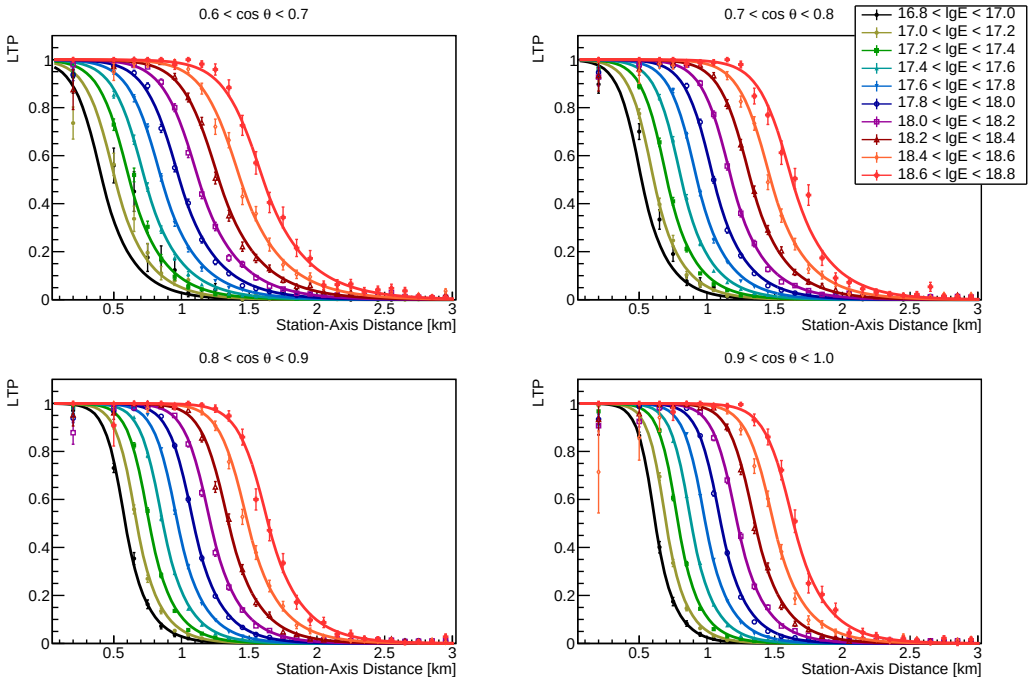


Figure A.20. The measured LTP data compared with the final parametrisation for ToT || TH-T1 || TH-T2 triggered stations.

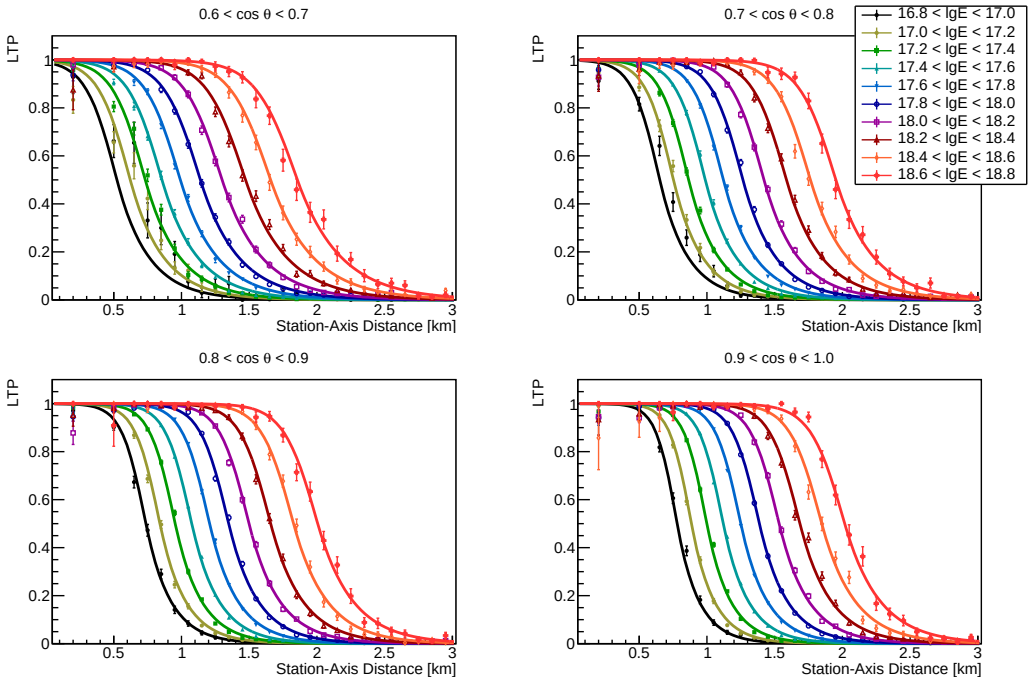


Figure A.21. The measured LTP data compared with the final parametrisation for ToT || TH-T1 || TH-T2 || ToTd || MoPS triggered stations.

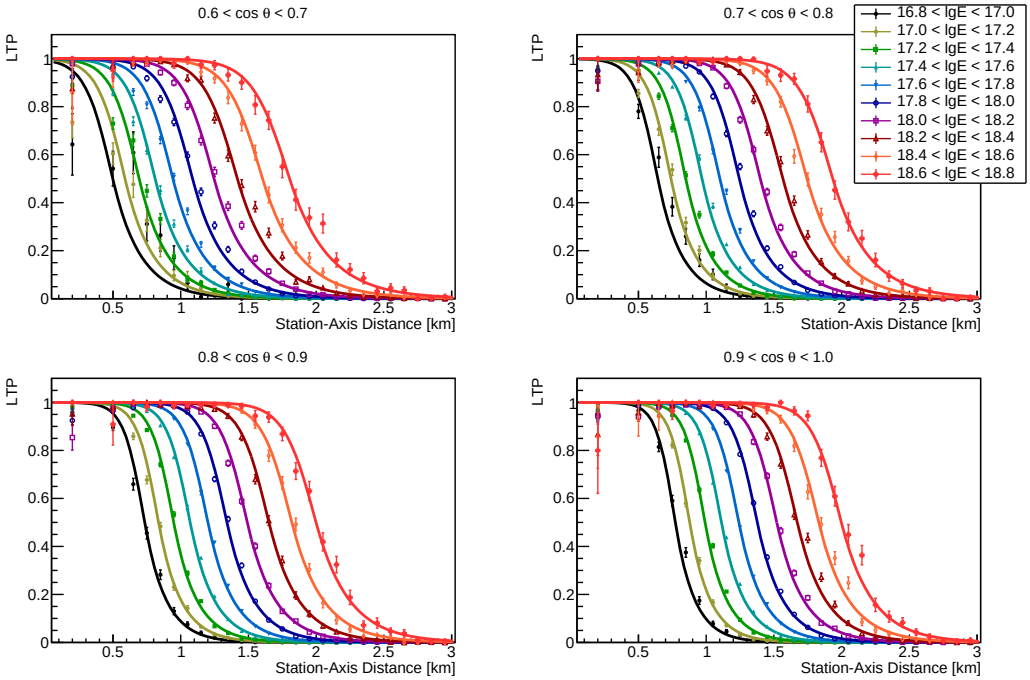


Figure A.22. The measured LTP data compared with the final parametrisation for ToT || ToTd || MoPS triggered stations.

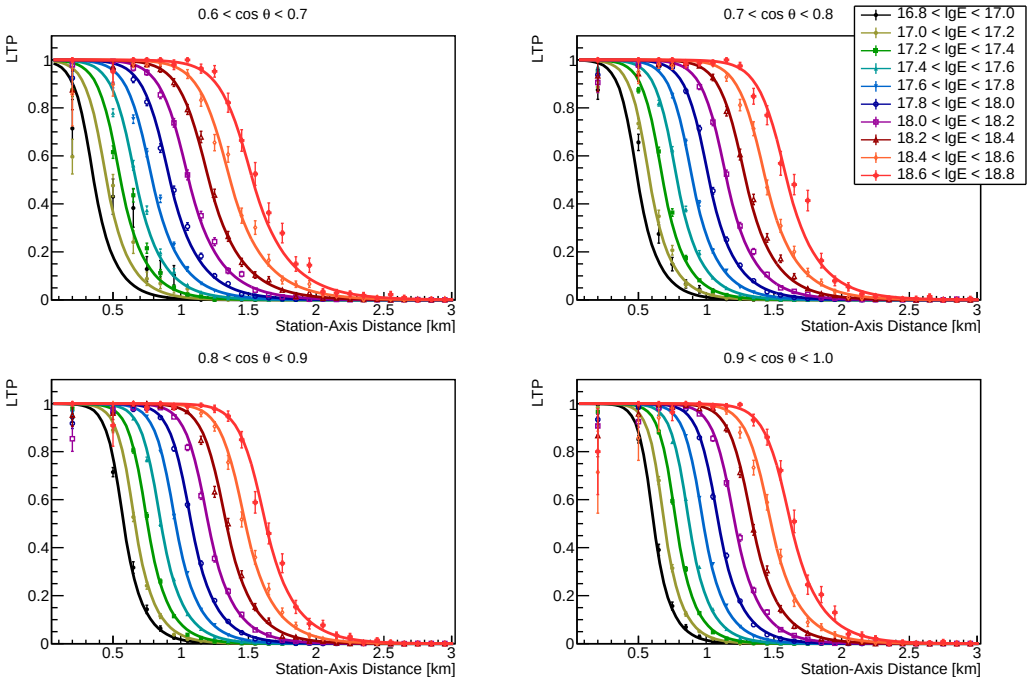


Figure A.23. The measured LTP data compared with the final parametrisation for ToT || TH-T1 || TH-T2 triggered stations.

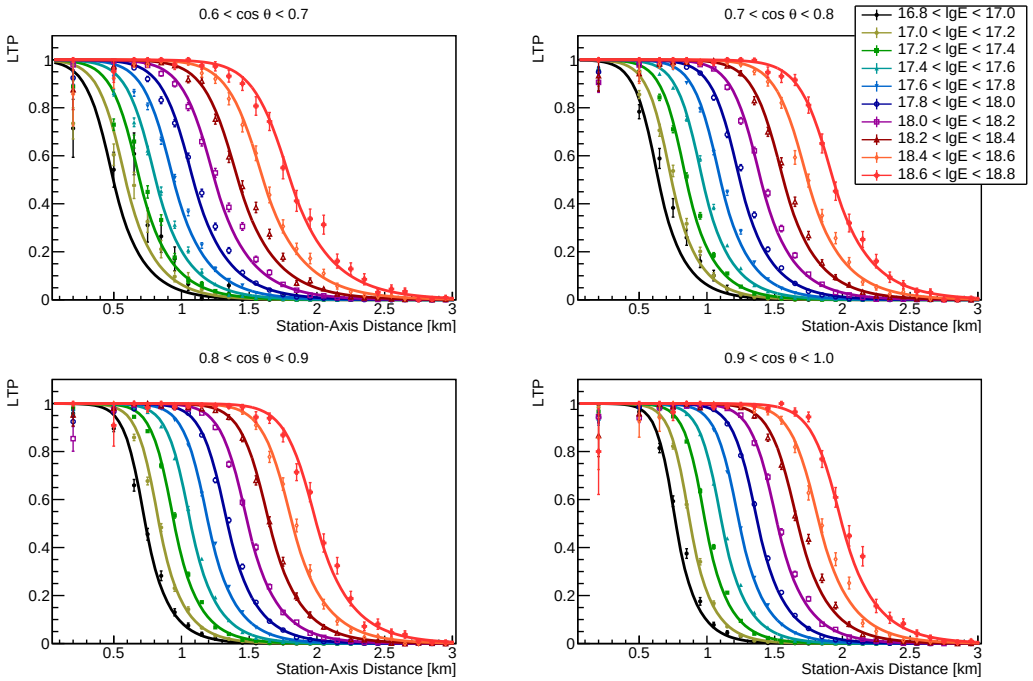


Figure A.24. The measured LTP data compared with the final parametrisation for ToT || TH-T2 || ToTd || MoPS triggered stations.

B

FAST Data Structure and Analysis Software

The Fluorescence detector Array of Single-pixel Telescopes (FAST) collaboration requires a convenient file format to store measured, simulated, and reconstructed data. Presented in this Appendix is a description of the FAST event structure which has been developed for the top-down reconstruction software. The event structure utilises the `ROOT` file format to allow simple handling of the complex data structures used by the FAST collaboration, and follows an object-oriented approach for storing data. The event structure is used to store the output of the detector simulation and reconstruction, and provides measured data in a convenient format for analysis. The event structure and software is available to FAST collaboration member through a `git` repository. The FAST data structure is inspired by the Pierre Auger Observatory (Auger) Offline and ADST data format. The basic structure of the data files is explained in the following sections.

B.1 File System and Event Structure

In the following sections, an overview of the data objects store in the event structure is presented. Brief examples of how to retrieve particular data is also given, however, more information is provided in the `doxygen` generated documentation.

B.1.1 FASTEventFile

The data files are stored as `TFiles`, making standard `ROOT` tools, such as `TBrowser`, available for inspecting the data. The `FASTEventFile` class provides an interface to read and write the data files in compiled programs. The user is not required to use any of the `ROOT` I/O methods to access the data. To read a `FASTEventFile`, the user can call

```
FASTEventFile eventFile("input.root", FASTEventFile::eRead);
```

This creates a `FASTEventFile` object called `eventFile` which contains the data from the “input.root” file.

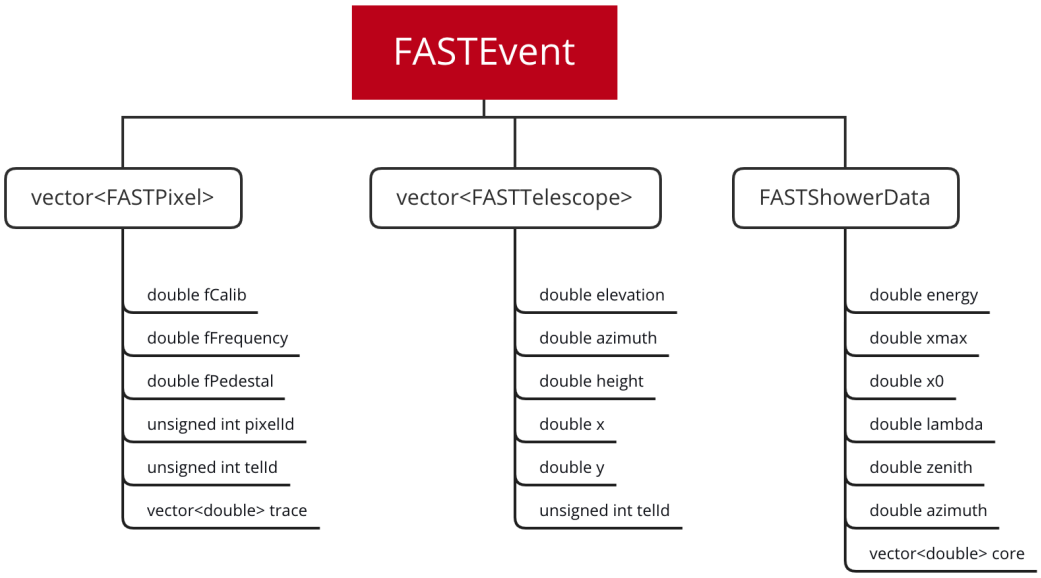


Figure B.1. Schematic of the `FASTEvent` classes and member variables.

B.1.2 FASTEvent

In order to access event data, the `FASTEvent` class is provided as an interface to the `TTree` stored within a `FASTEventFile`. The `FASTEvent` class contains important information related to an event, including event identification numbers, simulated and reconstructed shower parameters, detector configuration, and the individual photomultiplier tube (PMT) traces. The structure of the `FASTEvent` class is shown schematically in Figure B.1. Brief descriptions of the `FASTEvent` classes are provided below.

FASTPixel

The `FASTPixel` class contains pixel specific information, including the individual PMT traces, calibration information, as well as pixel and telescope identification numbers. Methods are provided to extract and set the values of the member variables. For example, the user can extract the trace from the ith pixel with the following function calls

```
vector<FASTPixel> pixels = event->GetPixels();
vector<double> trace = pixels[i].GetTrace();
```

FASTTelescope

The `FASTTelescope` class contains information about the detector state at the individual telescope level. This included the coordinates of the telescopes and their

pointing directions. The telescope identification numbers are also included so that information from different classes (e.g. `FASTPixel`) can be combined.

FASTShowerData

The `FASTShowerData` class contains the shower parameters such as energy, X_{max}, and shower geometry. The `FASTEvent` class contains two such objects; one for reconstructed, and one for simulated shower parameters. Such a format is extremely important for the top-down reconstruction, where measured events are compared to simulated events. This format also provides and convenient structure for reconstruction performance studies. Individual values can easily be extracted by the appropriate function calls. For example, to extract the reconstructed X_{max} value the user should call

```
double Xmax = event->GetShowerRecData().GetXmax();
```

Further information can be found in the detailed documentation provided with the FAST collaboration git repository. As the FAST experiment continues to grow and expand, this event structure will likely need to be adapted. The software has been developed with this in mind. Additional information, such as atmospheric conditions, will need to be added to the FAST data structure in the future.

References

[1] A. H. Bequerel. “On radioactivity, a new property of matter” (1903) (cit. on p. 5).

[2] C. A. de Coulomb. “Troisiéme Mémoire sur l’Electricité et le Magnétism”. *Histoire de l’Académie Royal des Sciences* (1785), 612 (cit. on p. 5).

[3] M. Faraday. “Experimental Researches in Electricity”. *Number v.2 in Experimental Researches in Electricity. Richard and John Edward Taylor, printers and publishers to the University of London* (1844) (cit. on p. 5).

[4] C. T. R. Wilson. “On the Ionisation of Atmospheric Air”. *Proceedings of the Royal Society of London* 68 (1901), 151–161 (cit. on p. 5).

[5] T. Wulf. “Beobachtungen überber Strahlung hoher Durchdringungsfäberhigkeit auf dem Eiffelturm”. *Physikalische Zeitschrift* 11 (1901), 811–813 (cit. on p. 6).

[6] V. F. Hess. “Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten / Observation of Penetrating Radiation in Seven Balloon Flights”. *Physikalische Zeitschrift* 13.13 (1912), 1084–1091 (cit. on p. 6).

[7] V. F. Hess. “Unsolved Problems in Physics: Tasks for the Immediate Future in Cosmic Ray Studies” (1936) (cit. on p. 6).

[8] J. Abraham et al. “Measurement of the energy spectrum of cosmic rays above 10^{18} eV using the Pierre Auger Observatory”. *Phys.Lett.* B685 (2010), 239–246. arXiv: 1002.1975 [astro-ph.HE] (cit. on pp. 6, 7).

[9] T. Abu-Zayyad et al. “The Cosmic Ray Energy Spectrum Observed with the Surface Detector of the Telescope Array Experiment”. *Astrophys.J.* 768 (2013), L1. arXiv: 1205.5067 [astro-ph.HE] (cit. on pp. 6, 43).

[10] R. U. Abbasi et al. “The energy spectrum of cosmic rays above $10^{17.2}$ eV measured by the fluorescence detectors of the Telescope Array experiment in seven Years”. *Astropart. Phys.* 80 (2016), 131–140. arXiv: 1511.07510 [astro-ph.HE] (cit. on p. 6).

[11] A. Aab et al. “The Pierre Auger Observatory: Contributions to the 33rd International Cosmic Ray Conference (ICRC 2013)” (2013). arXiv: 1307.5059 [astro-ph.HE] (cit. on p. 6).

[12] Blumer, J. and Engel, R. and Horandel, J. R. “Cosmic rays from the knee to the highest energies”. *Progress in Particle and Nuclear Physics* 63 (2009), 293–338. arXiv: 0904.0725v1 [astro-ph.HE] (cit. on pp. 6–8).

[13] F. G. Schröder. “News from Cosmic Ray Air Showers (ICRC 2019 – Cosmic Ray Indirect Rapport)”. *Proc. of the 36th International Cosmic Ray Conference, Madison, WI, USA* (2019). arXiv: 1910.03721v1 [astro-ph.HE] (cit. on p. 6).

[14] M. Nagano and A. A. Watson. “Observations and implications of the ultrahigh-energy cosmic rays”. *Rev. Mod. Phys.* 72 (3 July 2000), 689–732 (cit. on pp. 6, 26, 35, 38).

[15] J. R. Hörandel. “Models of the knee in the energy spectrum of cosmic rays”. *Astroparticle Physics* 21.3 (June 2004), 241–265 (cit. on p. 7).

[16] T. Antoni, W.D. Apel, A.F. Badea, et al. “KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems”. *Astroparticle Physics* 24.1 (2005), 1–25 (cit. on p. 7).

[17] V. S. Ptuskin, S. I. Rogovaya, V. N. Zirakashvili, et al. “Diffusion and drift of very high energy cosmic rays in galactic magnetic fields”. *aap* 268.2 (Feb. 1993), 726–735 (cit. on p. 7).

[18] P Sokolsky and G B Thomson. “Highest energy cosmic-rays and results from the HiRes experiment”. *Journal of Physics G: Nuclear and Particle Physics* 34.11 (Sept. 2007), R401–R429 (cit. on p. 7).

[19] W. D. Apel, J. C. Arteaga-Velázquez, K. Bekk, et al. “Kneelike Structure in the Spectrum of the Heavy Component of Cosmic Rays Observed with KASCADE-Grande”. *Phys. Rev. Lett.* 107 (17 Oct. 2011), 171104 (cit. on p. 7).

[20] D. Kang. “Latest results from the KASCADE-Grande data analysis”. *Proc. of the 36th International Cosmic Ray Conference, Madison, WI, USA* (2019), 306 (cit. on p. 7).

[21] A. Coleman. “Measurement of the Cosmic Ray Flux near the Second Knee with the Pierre Auger Observatory”. *Proc. of the 36th International Cosmic Ray Conference, Madison, WI, USA* (2019), 2335 (cit. on p. 7).

[22] “Measurement of the energy spectrum of ultra-high energy cosmic rays using the Pierre Auger Observatory”. *Proc. of the 36th International Cosmic Ray Conference, Madison, WI, USA* (2019), 450 (cit. on pp. 7, 8, 45, 70, 71).

[23] T. Abu-Zayyad, D. Ivanov, C. C. H. Jui, et al. “The Knee and the Second Knee of the Cosmic-Ray Energy Spectrum” (2018). arXiv: 1803.07052 [astro-ph.HE] (cit. on p. 7).

[24] Roberto Aloisio, Veniamin Berezinsky, Pasquale Blasi, et al. “A dip in the UHECR spectrum and the transition from galactic to extragalactic cosmic rays”. *Astroparticle Physics* 27.1 (2007), 76–91 (cit. on p. 7).

[25] Veniamin Berezinsky, Askhat Gazizov, and Svetlana Grigorieva. “On astrophysical solution to ultrahigh energy cosmic rays”. *Phys. Rev. D* 74 (4 Aug. 2006), 043005 (cit. on p. 7).

[26] T. Abu-Zayyad et al. “Measurement of the cosmic ray energy spectrum and composition from 10^{17} eV to $10^{18.3}$ eV using a hybrid fluorescence technique”. *Astrophys.J.* 557 (2001), 686–699. arXiv: astro-ph/0010652 [astro-ph] (cit. on p. 7).

[27] The Pierre Augur collaboration. “Measurement of the cosmic ray spectrum above 4×1018eV using inclined events detected with the Pierre Auger Observatory”. *Journal of Cosmology and Astroparticle Physics* 2015.08 (Aug. 2015), 049–049 (cit. on p. 7).

[28] O. Deligny. “The energy spectrum of ultra-high energy cosmic rays measured at the Pierre Auger Observatory and at the Telescope Array”. *Proc. of the 36th International Cosmic Ray Conference, Madison, WI, USA* (2019), 234 (cit. on pp. 8, 297).

[29] R.U. Abbasi, T. Abu-Zayyad, J.F. Amman, et al. “Observation of the ankle and evidence for a high-energy break in the cosmic ray spectrum”. *Physics Letters B* 619.3 (2005), 271–280 (cit. on p. 8).

[30] R. U. Abbasi, T. Abu-Zayyad, M. Allen, et al. “First Observation of the Greisen-Zatsepin-Kuzmin Suppression”. *Physical Review Letters* 100.10 (Mar. 2008) (cit. on p. 8).

[31] K. Greisen. “End to the cosmic ray spectrum?” *Phys. Rev. Lett.* 16 (1966), 748–750 (cit. on p. 8).

[32] G.T. Zatsepin and V.A. Kuzmin. “Upper limit of the spectrum of cosmic rays”. *JETP Lett.* 4 (1966), 78–80 (cit. on p. 8).

[33] A.A. Penzias and R.W. Wilson. “A Measurement of Excess Antenna Temperature at 4080 Mc/s”. *Astrophys. J.* 142 (1965), 419–421 (cit. on p. 8).

[34] G.R. Blumenthal. “Energy Loss of High-Energy Cosmic Rays in Pair- Producing Collisions with Ambient Photons”. *Physical Review D* 1 (1970), 1596–1602 (cit. on p. 9).

[35] J.L. Puget, F.W. Stecker, and J.H. Bredekamp. “Photonuclear interactions of ultra-high energy cosmic rays and their astrophysical consequences”. *Astrophys. J.* 205 (1976), 638–654 (cit. on p. 9).

[36] J.W. Cronin. “Cosmic rays: the most energetic particles in the universe”. *Reviews of Modern Physics Supplement* 71 (1999), 165 (cit. on p. 9).

[37] E. Roulet. “Large-scale anisotropies above 0.03 EeV measured by the Pierre Auger Observatory”. *Proc. of the 36th International Cosmic Ray Conference, Madison, WI, USA* (2019), 408 (cit. on pp. 9, 17, 18).

[38] A. Aab et al. “Observation of a Large-scale Anisotropy in the Arrival Directions of Cosmic Rays above 8×10^{18} eV”. *Science* 357.6537 (2017), 1266–1270. arXiv: 1709.07321 [astro-ph.HE] (cit. on pp. 9, 17).

[39] P. Bhattacharjee and G. Sigl. “Origin and propagation of extremely high-energy cosmic rays”. *Physics Reports* 327.3-4 (Apr. 2000), 109–247 (cit. on p. 10).

[40] D. Hooper, A. M. Taylor, and S. Sarkar. “Cosmogenic photons as a test of ultra-high energy cosmic ray composition”. *Astroparticle Physics* 34.6 (Jan. 2011), 340–343 (cit. on p. 10).

[41] O. E. Kalashev, V. A. Kuzmin, and D. V. Semikoz. “Top-Down Models and Extremely High Energy Cosmic Rays” (1999). arXiv: astro-ph/9911035 [astro-ph] (cit. on p. 10).

[42] J. Abraham, P. Abreu, M. Aglietta, et al. “Upper limit on the cosmic-ray photon flux above 10^{19} eV using the surface detector of the Pierre Auger Observatory”. *Astroparticle Physics* 29.4 (May 2008), 243–256 (cit. on p. 11).

[43] J. Abraham, P. Abreu, M. Aglietta, et al. “Upper limit on the cosmic-ray photon fraction at EeV energies from the Pierre Auger Observatory”. *Astroparticle Physics* 31.6 (July 2009), 399–406 (cit. on p. 11).

[44] S. Petrera. “Photons in the science of the Pierre Auger Observatory” (2019). arXiv: 1910.02704 [astro-ph.CO] (cit. on p. 11).

[45] E. Fermi. “On the Origin of the Cosmic Radiation”. *Phys. Rev.* 75 (8 Apr. 1949), 1169–1174 (cit. on p. 11).

[46] R. D. Blandford and J. P. Ostriker. “Particle acceleration by astrophysical shocks.” *The Astrophysical Journal* 221 (Apr. 1978), L29–L32 (cit. on pp. 11, 12).

[47] R. J. Protheroe. “Acceleration and Interaction of Ultra High Energy Cosmic Rays” (1998). arXiv: astro-ph/9812055 [astro-ph] (cit. on pp. 11–14).

[48] W. I. Axford, E. Leer, and G. Skadron. “The Acceleration of Cosmic Rays by Shock Waves”. International Cosmic Ray Conference 11 (Jan. 1977), 132 (cit. on p. 12).

[49] A. R. Bell. “The acceleration of cosmic rays in shock fronts – I”. *Monthly Notices of the Royal Astronomical Society* 182.2 (Feb. 1978), 147–156. eprint: https://academic.oup.com/mnras/article-pdf/182/2/147/3710138/mnras182-0147.pdf (cit. on p. 12).

[50] A. R. Bell. “The acceleration of cosmic rays in shock fronts – II”. *Monthly Notices of the Royal Astronomical Society* 182.3 (Mar. 1978), 443–455. eprint: https://academic.oup.com/mnras/article-pdf/182/3/443/3856040/mnras182-0443.pdf (cit. on p. 12).

[51] T. Stanev. *Ultra-High-energy Cosmic Rays: Origin and Propagation*. Ed. by J Wefel, T Stanev, and M Shapiro. World Scientific Publishing Co. Pte. Ltd., 2007, 83–96 (cit. on p. 15).

[52] A. M. Hillas. “The Origin of Ultrahigh-Energy Cosmic Rays”. *Ann. Rev. Astron. Astrophys.* 22 (1984), 425–444 (cit. on pp. 15, 16).

[53] Diego F Torres and Luis A Anchordoqui. “Astrophysical origins of ultrahigh energy cosmic rays”. *Reports on Progress in Physics* 67.9 (July 2004), 1663–1730 (cit. on p. 15).

[54] Rainer Beck. “Galactic and Extragalactic Magnetic Fields”. *The Astrophysics of Galactic Cosmic Rays*. Ed. by Roland Diehl, Etienne Parizot, Reinald Kallenbach, and Rudolf Von Steiger. Dordrecht: Springer Netherlands, 2001, 243–260 (cit. on pp. 15, 16).

[55] P. P. Kronberg. “Extragalactic magnetic fields”. *Reports on Progress in Physics* 57.4 (Apr. 1994), 325–382 (cit. on p. 15).

[56] Felix Aharonian, Andrei Bykov, Etienne Parizot, Vladimir Ptuskin, and Alan Watson. “Cosmic Rays in Galactic and Extragalactic Magnetic Fields”. *Space Science Reviews* 166.1-4 (Apr. 2011), 97–132 (cit. on p. 16).

[57] A. Aab et al. “Large-scale cosmic-ray anisotropies above 4 EeV measured by the Pierre Auger Observatory”. *Astrophys. J.* 868.1 (2018), 4. arXiv: 1808.03579 [astro-ph.HE] (cit. on p. 17).

[58] R. Abbasi, Y. Abdou, T. Abu-Zayyad, et al. “Observation of Anisotropy in the Galactic Cosmic-Ray Arrival Directions at 400 TeV with IceCube”. *The Astrophysical Journal* 746.1 (Jan. 2012), 33 (cit. on pp. 17, 18).

[59] M. G. Aartsen, K. Abraham, M. Ackermann, et al. “Anisotropy in Cosmic-Ray Arrival Directions in the Southern Hemipshere Based on Six Years of Data from the IceCube Detector”. *The Astrophysical Journal* 826.2 (Aug. 2016), 220 (cit. on pp. 17, 18).

[60] A. Chiavassa, W.D. Apel, J.C. Arteaga-Velázquez, et al. “KASCADE-Grande experiment measurements of the cosmic ray spectrum and large scale anisotropy”. *Nuclear and Particle Physics Proceedings* 279-281 (2016). Proceedings of the 9th Cosmic Ray International Seminar, 56–62 (cit. on pp. 17, 18).

[61] L. Caccianiga. “Anisotropies of the highest energy cosmic-ray events recorded by the Pierre Auger Observatory in 15 years of operation”. *Proc. of the 36th International Cosmic Ray Conference, Madison, WI, USA* (2019), 206 (cit. on pp. 17, 18).

[62] A. Aab et al. “An Indication of anisotropy in arrival directions of ultra-high-energy cosmic rays through comparison to the flux pattern of extragalactic gamma-ray sources”. *Astrophys. J.* 853.2 (2018), L29. arXiv: 1801.06160 [astro-ph.HE] (cit. on p. 17).

[63] Armando di Matteo, Toshihiro Fujii, and Kazumasa Kawata. “Investigating an angular correlation between nearby starburst galaxies and UHECRs with the Telescope Array experiment”. *EPJ Web of Conferences* 210 (2019). Ed. by I. Lhenry-Yvon, J. Biteau, O. Deligny, and P.Editors Ghia, 01007 (cit. on p. 17).

[64] R.U. Abbasi et al. “Indications of Intermediate-Scale Anisotropy of Cosmic Rays with Energy Greater Than 57 EeV in the Northern Sky Measured with the Surface Detector of the Telescope Array Experiment”. *Astrophys.J.* 790 (2014), L21. arXiv: 1404.5890 [astro-ph.HE] (cit. on pp. 17, 18, 43).

[65] K. Kawata. “Updated Results on the UHECR Hotspot Observed by the Telescope Array Experiment”. *PoS* ICRC2019 (2019), 310 (cit. on pp. 17, 18, 43).

[66] P. Abreu et al. “Update on the correlation of the highest energy cosmic rays with nearby extragalactic matter”. *Astropart.Phys.* 34 (2010), 314–326. arXiv: 1009.1855 [astro-ph.HE] (cit. on p. 18).

[67] A. di Matteo and others. “Full-sky searches for anisotropies in UHECR arrival directions with the Pierre Auger Observatory and the Telescope Array”. *PoS* ICRC2019 (2019), 439. arXiv: 2001.01864 [astro-ph.HE] (cit. on p. 18).

[68] P. Auger, P. Ehrenfest, R. Maze, J. Daudin, and R. A. Fréon. “Extensive Cosmic-Ray Showers”. *Rev. Mod. Phys.* 11 (3-4 July 1939), 288–291 (cit. on p. 19).

[69] Andreas Haungs, Heinigerd Rebel, and Markus Roth. “Energy spectrum and mass composition of high-energy cosmic rays”. *Reports on Progress in Physics* 66.7 (June 2003), 1145–1206 (cit. on p. 20).

[70] W. Heitler. *The Quantum Theory of Radiation*. 3rd ed. Oxford University Press, 1954 (cit. on p. 21).

[71] Paul Sommers. “Extensive air showers and measurement techniques”. *Comptes Rendus Physique* 5.4 (2004). Ultimate energy particles in the Universe, 463–472 (cit. on p. 21).

[72] J. Matthews. “A Heitler model of extensive air showers”. *Astroparticle Physics* 22 (2005), 387–397 (cit. on pp. 21–25).

[73] P. Abreu, M. Aglietta, E. J. Ahn, et al. “Measurement of the Proton-Air Cross Section at sqrt(s)=57 TeV with the Pierre Auger Observatory”. *Physical Review Letters* 109.6 (Aug. 2012) (cit. on p. 22).

[74] Tim Huege and Dave Besson. “Radio-wave detection of ultra-high-energy neutrinos and cosmic rays”. *Progress of Theoretical and Experimental Physics* 2017.12 (Nov. 2017) (cit. on p. 26).

[75] K.H. Kampert and M. Unger. “Measurements of the cosmic ray composition with air shower experiments”. *Astroparticle Physics* 35.10 (May 2012), 660–678 (cit. on p. 27).

[76] L. Cazon. “Probing High-Energy Hadronic Interactions with Extensive Air Showers”. *Proc. of the 36th International Cosmic Ray Conference, Madison, WI, USA* (2019), 005. arXiv: 1909.02962 [hep-ex] (cit. on p. 28).

[77] S. Ostapchenko. “Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: QGSJET-II model”. *Physical Review D* 83.1 (Jan. 2011) (cit. on p. 28).

[78] Ostapchenko, S. “QGSJET-II: physics, recent improvements, and results for air showers”. *EPJ Web of Conferences* 52 (2013), 02001 (cit. on p. 28).

[79] T. Pierog, Iu. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner. “EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider”. *Physical Review C* 92.3 (Sept. 2015) (cit. on p. 28).

[80] F. Riehn, R. Engel, A. Fedynitch, T.K. Gaisser, and T. Stanev. “A new version of the event generator Sibyll” (2015). arXiv: 1510.00568 [hep-ph] (cit. on p. 28).

[81] R. Engel, A. Fedynitch, T.K. Gaisser, F. Riehn, and T. Stanev. “The hadronic interaction model Sibyll 2.3c and extensive air showers” (2019). arXiv: 1912.03300 [hep-ph] (cit. on pp. 28, 29, 136).

[82] L. Nellen. “The observation of a muon deficit in simulations from data of the Pierre Auger Observatory”. *Journal of Physics: Conference Series* 409 (Feb. 2013), 012107 (cit. on p. 28).

[83] H.P. Dembinski, J.C. Arteaga-Velázquez, L. Cazon, et al. “Report on Tests and Measurements of Hadronic Interaction Properties with Air Showers”. *EPJ Web of Conferences* 210 (2019). Ed. by I. Lhenry-Yvon, J. Biteau, O. Deligny, and P.Editors Ghia, 02004 (cit. on p. 28).

[84] L. Cazon. “Working Group Report on the Combined Analysis of Muon Density Measurements from Eight Air Shower Experiments”. *Proc. of the 36th International Cosmic Ray Conference, Madison, WI, USA* (2019), 214 (cit. on p. 28).

[85] A. Aab, P. Abreu, M. Aglietta, et al. “Testing Hadronic Interactions at Ultrahigh Energies with Air Showers Measured by the Pierre Auger Observatory”. *Phys. Rev. Lett.* 117 (19 Oct. 2016), 192001 (cit. on p. 28).

[86] V. de Souza. “Testing the agreement between the $X_{\max}$ distributions measured by the Pierre Auger and Telescope Array Observatories”. *PoS* ICRC2017 (2018), 522 (cit. on p. 29).

[87] E. Barcikowski et al. “Mass Composition Working Group Report at UHECR-2012”. *EPJ Web Conf.* 53 (2013), 01006. arXiv: 1306.4430 [astro-ph.HE] (cit. on p. 29).

[88] R. Abbasi et al. “Report of the Working Group on the Composition of Ultra High Energy Cosmic Rays”. *Proc. of 2014 Conference on Ultrahigh Energy Cosmic Rays, USA, Utah (2014)* (2015). arXiv: 1503.07540 [astro-ph.HE] (cit. on p. 29).

[89] Yushkov, A., Bellido, J., Belz, J., et al. “Depth of maximum of air-shower profiles: testing the compatibility of measurements performed at the Pierre Auger Observatory and the Telescope Array experiment”. *EPJ Web Conf.* 210 (2019), 01009 (cit. on pp. 29–31).

[90] A. Yushkov. “Mass Composition of Cosmic Rays with Energies above 1017.2 eV from the Hybrid Data of the Pierre Auger Observatory”. *Proc. of the 36th International Cosmic Ray Conference, Madison, WI, USA* (2019), 482 (cit. on pp. 29, 30, 32, 172, 182–184).

[91] W. Hanlon. “Telescope Array 10 Year Composition” (2019). arXiv: 1908.01356 [astro-ph.HE] (cit. on pp. 30, 31).

[92] P. Abreu et al. *Journal of Cosmology and Astroparticle Physics* 2013.02 (Feb. 2013), 026–026 (cit. on p. 32).

[93] J. Bellido et al. “Depth of maximum of air-shower profiles at the Pierre Auger Observatory: Measurements above $10^{17.2}$ eV and Composition Implications”. *Proc. of the 35th International Cosmic Ray Conference* 301 (2017), 506 (cit. on pp. 32, 85).

[94] C. J. Todero Peixoto. “Estimating the Depth of Shower Maximum using the Surface Detectors of the Pierre Auger Observatory”. *Proc. of the 36th International Cosmic Ray Conference, Madison, WI, USA* (2019), 440 (cit. on pp. 32–34).

[95] A. Aab et al. “Inferences on mass composition and tests of hadronic interactions from 0.3 to 100 EeV using the water-Cherenkov detectors of the Pierre Auger Observatory”. *Phys. Rev.* D96.12 (2017), 122003. arXiv: 1710.07249 [astro-ph.HE] (cit. on pp. 32, 34).

[96] R.U. Abbasi et al. “Indications of Proton-Dominated Cosmic Ray Composition above 1.6 EeV”. *Phys.Rev.Lett.* 104 (2010), 161101. arXiv: 0910.4184 [astro-ph.HE] (cit. on pp. 34, 40).

[97] A. Aab et al. “Depth of maximum of air-shower profiles at the Pierre Auger Observatory. I. Measurements at energies above $10^{17.8}$ eV”. *Phys.Rev.* D90.12 (2014), 122005. arXiv: 1409.4809 [astro-ph.HE] (cit. on pp. 35, 85, 111–114).

[98] A. Aab et al. “Depth of maximum of air-shower profiles at the Pierre Auger Observatory. II. Composition implications”. *Phys.Rev.* D90.12 (2014), 122006. arXiv: 1409.5083 [astro-ph.HE] (cit. on p. 35).

[99] J. Linsley, L. Scarsi, and B. Rossi. “Extremely Energetic Cosmic-Ray Event”. *Phys. Rev. Lett.* 6 (9 May 1961), 485–487 (cit. on p. 35).

[100] L. Bonolis. “From cosmic ray physics to cosmic ray astronomy: Bruno Rossi and the opening of new windows on the universe” (2012). arXiv: 1211.4061 [physics.hist-ph] (cit. on p. 35).

[101] J. Linsley, J L. Scarsi, and B. Rossi. “Energy Spectrum and Structure of Large Air Showers”. *Journal of the Physical Society Japan Supplement A-III* 17 (1962), 91–102 (cit. on pp. 35, 36).

[102] J. Linsley. “Evidence for a Primary Cosmic-Ray Particle with Energy 10^{20} eV”. *Phys. Rev. Lett.* 10 (4 Feb. 1963), 146–148 (cit. on pp. 35, 36).

[103] J. Linsley. “Primary cosmic rays of energy 10^{17} to 10^{20} eV, the energy spectrum and arrival directions”. *Proc. of 8th Int. Cosmic Ray Conf* 4 (1963), 77–99 (cit. on p. 35).

[104] H. E. Bergeson, G. L. Cassiday, T. -W. Chiu, et al. “Measurement of Light Emission from Remote Cosmic-Ray Air Showers”. *Phys. Rev. Lett.* 39 (13 Sept. 1977), 847–849 (cit. on p. 35).

[105] P. Blake, H. Ferguson, and W. Nash. “Current Studies of Extensive Air Showers by the Nottingham Group at Haverah Park”. *Journal of The Physical Society of Japan - J PHYS SOC JPN* 33 (Nov. 1972) (cit. on pp. 36, 37).

[106] M. A. Lawrence, R. J. O. Reid, and A. A. Watson. “The cosmic ray energy spectrum above 4x1017eV as measured by the Haverah Park array”. *Journal of Physics G: Nuclear and Particle Physics* 17.5 (May 1991), 733–757 (cit. on p. 36).

[107] R. M. Tennent. “The Haverah Park extensive air shower array”. *Proceedings of the Physical Society* 92.3 (Nov. 1967), 622–631 (cit. on p. 36).

[108] A. J. Bower, G. Brooke, D. Pearce, J. C. Perrett, and A. A. Watson. “The lateral distribution of scintillator densities in air showers produced by cosmic rays with E > 5x1018eV”. *Journal of Physics G: Nuclear Physics* 9.12 (Dec. 1983), 1569–1576 (cit. on p. 36).

[109] L.J. Kewley, R.W. Clay, and B.R. Dawson. “Arrival directions of the southern highest energy cosmic rays”. *Astroparticle Physics* 5.1 (1996), 69–74 (cit. on p. 38).

[110] R. G. Brownlee, A. J. Fisher, L. Goorevich, et al. “Design of an array to record air showers of energy up to 1021 eV”. *Canadian Journal of Physics* 46.10 (1968), S259–S262. eprint: https://doi.org/10.1139/p68-223 (cit. on p. 38).

[111] C. J. Bell. “The calibration of large-area liquid scintillation counters used in a cosmic ray experiment”. *Journal of Physics G: Nuclear Physics* 2.11 (Nov. 1976), 857–866 (cit. on p. 38).

[112] Anatoly, Ivanov. “The Yakutsk array experiment: Main results and future directions”. *EPJ Web of Conferences* 53 (2013), 04003 (cit. on p. 38).

[113] A. A. Ivanov, S. P. Knurenko, and I. Ye Sleptsov. “Measuring extensive air showers with Cherenkov light detectors of the Yakutsk array: the energy spectrum of cosmic rays”. *New Journal of Physics* 11.6 (June 2009), 065008 (cit. on pp. 38, 39).

[114] Ivanov, Anatoly. “Some results from the Yakutsk array experiment and the status of the modernization program”. *EPJ Web Conf.* 208 (2019), 08005 (cit. on p. 38).

[115] R.M. Baltrusaitis, R. Cady, G.L. Cassiday, et al. “The Utah Fly’s Eye detector”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 240.2 (1985), 410–428 (cit. on p. 39).

[116] D. J. Bird, S. C. Corbato, H. Y. Dai, et al. “Results from the fly’s eye experiment”. *AIP Conference Proceedings* 338.1 (1995), 839–854. eprint: https://aip.scitation.org/doi/pdf/10.1063/1.48457 (cit. on p. 39).

[117] CERN Timelines. *The Fly’s Eye Mirrors*. 1991. URL: http://timeline-fr.web.cern.ch/fly%E2%80%99s-eye-detects-record-breaking-cosmic-ray-32-x-1020-ev (cit. on p. 40).

[118] Pierre Sokolsky, Paul Sommers, and Bruce R. Dawson. “Extremely high energy cosmic rays”. *Physics Reports* 217.5 (1992), 225–277 (cit. on pp. 39, 40).

[119] T Abu-Zayyad, M Al-Seady, K Belov, et al. “The prototype high-resolution Fly’s Eye cosmic ray detector”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 450.2 (2000), 253–269 (cit. on p. 40).

[120] J.H. Boyer, B.C. Knapp, E.J. Mannel, and M. Seman. “FADC-based DAQ for HiRes Fly’s Eye”. *Nucl. Instrum. Meth.* A482 (2002), 457–474 (cit. on p. 40).

[121] R.U. Abbasi et al. “First observation of the Greisen-Zatsepin-Kuzmin suppression”. *Phys.Rev.Lett.* 100 (2008), 101101. arXiv: astro-ph/0703099 [astro-ph] (cit. on pp. 40, 41).

[122] P. Sokolsky. “Final Results from the High solution Fly’s Eye (HiRes) Experiment”. *Nuclear Physics B - Proceedings Supplements* 212-213 (Mar. 2011), 74–78 (cit. on p. 40).

[123] N. Chiba, K. Hashimoto, N. Hayashida, et al. “Akeno Giant Air Shower Array (AGASA) covering 100 km2 area”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 311.1 (1992), 338–349 (cit. on pp. 40, 41).

[124] Kenji Shinozaki and Masahiro Teshima. “AGASA Results”. *Nuclear Physics B - Proceedings Supplements* 136 (2004). CRIS 2004 Proceedings of the Cosmic Ray International Seminars: GZK and Surroundings, 18–27 (cit. on pp. 40, 41).

[125] Kenji Shinozaki. “AGASA results”. *Nuclear Physics B - Proceedings Supplements* 151.1 (2006). VERY HIGH ENERGY COSMIC RAY INTERACTIONS, 3–10 (cit. on pp. 40, 41).

[126] M. Takeda et al. “Extension of the cosmic ray energy spectrum beyond the predicted Greisen-Zatsepin-Kuz’min cutoff”. *Phys. Rev. Lett.* 81 (1998), 1163–1166. arXiv: astro-ph/9807193 [astro-ph] (cit. on p. 41).

[127] M. Fukushima and others. “Telescope array project for extremely high energy cosmic rays”. *Prog.Theor.Phys.Suppl.* 151 (2003), 206–210 (cit. on p. 41).

[128] H. Tokuno, Y. Tameda, M. Takeda, et al. “New air fluorescence detectors employed in the Telescope Array experiment”. *Nucl. Instrum. Meth.* A676 (2012), 54–65. arXiv: 1201.0002 [astro-ph.IM] (cit. on pp. 41, 42).

[129] T. Abu-Zayyad and others. “The surface detector array of the Telescope Array experiment”. *Nucl. Instrum. Meth.* A689 (2012), 87–97. arXiv: 1201.4964 [astro-ph.IM] (cit. on p. 41).

[130] T. Abu-Zayyad, R. Aida, M. Allen, et al. “The energy spectrum of ultra-high-energy cosmic rays measured by the Telescope Array FADC fluorescence detectors in monocular mode”. *Astroparticle Physics* 48 (2013), 16–24 (cit. on p. 42).

[131] S. Udo. “The Telescope Array Low-energy Extension”. *PoS* ICRC2017 (2017), 544 (cit. on pp. 42, 44).

[132] S. Ogio. “Telescope Array Low-energy Extension (TALE) hybrid”. *PoS* ICRC2019 (2019), 375 (cit. on pp. 42, 44).

[133] D. Ivanov. “Energy Spectrum Measured by the Telescope Array Experiment”. *PoS* ICRC2019 (2019), 298 (cit. on p. 43).

[134] E. Kido. “The TAx4 experiment”. *PoS* ICRC2017 (2017), 386 (cit. on p. 43).

[135] E. Kido. “Status and prospects of the TAx4 experiment”. *PoS* ICRC2019 (2019), 312 (cit. on pp. 43, 44).

[136] A. Aab et al. “The Pierre Auger Cosmic Ray Observatory”. *Nucl. Instrum. Meth.* A798 (2015), 172–213. arXiv: 1502.01323 [astro-ph.IM] (cit. on pp. 45–48, 51, 54–57, 60, 61, 64, 66–68, 73, 85, 187, 196, 199, 266).

[137] I. Allekotte et al. “The Surface Detector System of the Pierre Auger Observatory”. *Nucl. Instrum. Meth.* A586 (2008), 409–420. arXiv: 0712.2832 [astro-ph] (cit. on pp. 45, 47).

[138] Karl-Heinz Kampert and Alan A. Watson. “Extensive air showers and ultra high-energy cosmic rays: a historical review”. *The European Physical Journal H* 37.3 (July 2012), 359–412 (cit. on p. 48).

[139] X. Bertou, P.S. Allison, C. Bonifazi, et al. “Calibration of the surface array of the Pierre Auger Observatory”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 568.2 (2006), 839–846 (cit. on pp. 47–49).

[140] Pierre Auger Collaboration. “Trigger and aperture of the surface detector array of the Pierre Auger Observatory”. *Nuclear Instruments and Methods A* 613 (2010), 29–39. arXiv: 1111.6764v1 [astro-ph.IM] (cit. on pp. 49, 52, 53).

[141] A. Coleman. “Measurement of the Cosmic Ray Flux above 100 PeV at the Pierre Auger Observatory”. *PhD thesis, Pennsylvania State University* (2018) (cit. on pp. 51, 103).

[142] M. Ave et al. “Precise measurement of the absolute fluorescence yield of the 337 nm band in atmospheric gases”. *Astropart.Phys.* 42 (2013), 90–102. arXiv: 1210.6734 [astro-ph.IM] (cit. on p. 54).

[143] M. Ave et al. “Measurement of the pressure dependence of air fluorescence emission induced by electrons”. *Astropart.Phys.* 28 (2007), 41–57. arXiv: astro-ph/0703132 [ASTRO-PH] (cit. on p. 54).

[144] J. Abraham et al. “The Fluorescence Detector of the Pierre Auger Observatory”. *Nucl. Instrum. Meth.* A620 (2010), 227–251. arXiv: 0907.4282 [astro-ph.IM] (cit. on pp. 54–60, 74–76, 81, 187).

[145] H. Gemmeke, M. Kleifges, and A. Menshikov. “Statistical calibration and background measurements of the Auger fluorescence detector”. *Forschungszentrum Karlsruhe Report FZKA* (2003), 891–894 (cit. on p. 54).

[146] H. Gemmeke. “The Auger Fluorescence Detector Electronics”. *Proc. of the 27th International Cosmic Ray Conference, Hamburg, Germany* (2001), 737 (cit. on p. 56).

[147] A. Schmidt, T. Asch, H. Gemmeke, et al. “Third Level Trigger for the Fluorescence Telescopes of the Pierre Auger Observatory”. *Nucl. Instrum. Meth.* A601 (2009), 347–353. arXiv: 0807.4421 [astro-ph] (cit. on pp. 56, 57).

[148] A. Aab, P. Abreu, M. Aglietta, et al. “Spectral calibration of the fluorescence telescopes of the Pierre Auger Observatory”. *Astroparticle Physics* 95 (Oct. 2017), 44–56 (cit. on pp. 57–59).

[149] R. Knapik, P. Bauleo, B. R. Becker, et al. “The Absolute, Relative and Multi-Wavelength Calibration of the Pierre Auger Observatory Fluorescence Detectors” (2007). arXiv: 0708.1924 [astro-ph] (cit. on pp. 57–59).

[150] A.C. Rovero, P. Bauleo, J.T. Brack, J.L. Harton, and R. Knapik. “Multi-wavelength calibration procedure for the Pierre Auger Observatory Fluorescence Detectors”. *Astroparticle Physics* 31.4 (2009), 305–311 (cit. on pp. 57, 58).

[151] M. D. Roberts. “Calibration of the Pierre Auger fluorescence detector” (2003). arXiv: astro-ph/0308410 [astro-ph] (cit. on p. 57).

[152] C. Meurer and N. Scharf. “HEAT – a low energy enhancement of the Pierre Auger Observatory”. *Astrophysics and Space Sciences Transactions* 7.2 (May 2011), 183–186 (cit. on p. 60).

[153] B. Fick, M. Malek, J. A. J. Matthews, J. Matthews, R. Meyhandan, et al. “The Central Laser Facility at the Pierre Auger Observatory”. *J.Instrum.* 1.11 (2006), P11003–P11003 (cit. on p. 61).

[154] V. Harvey. “Real-time Measurements with Atmsopheric Instruments at the Pierre Auger Observatory”. *Proc. of the 36th International Cosmic Ray Conference, Madison, WI, USA* (2019), 283 (cit. on pp. 61–63).

[155] The Pierre Auger Collaboration. “Techniques for measuring aerosol attenuation using the Central Laser Facility at the Pierre Auger Observatory”. *Journal of Instrumentation* 8.04 (Apr. 2013), P04009–P04009 (cit. on pp. 61, 62).

[156] L. Valore. “Atmospheric aerosol characterization using the central laser facility at the Pierre Auger Observatory”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 630.1 (2011). Proceedings of the 2nd Roma International Conference on Astroparticle Physics (RICAP 2009), 246–250 (cit. on p. 61).

[157] J. Abraham, P. Abreu, M. Aglietta, et al. “A study of the effect of molecular and aerosol conditions in the atmosphere on air fluorescence measurements at the Pierre Auger Observatory”. *Astroparticle Physics* 33.2 (Mar. 2010), 108–129. arXiv: 1002.0366 [astro-ph.IM] (cit. on pp. 61, 257).

[158] S.Y. BenZvi, R. Cester, M. Chiosso, et al. “The Lidar system of the Pierre Auger Observatory”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 574.1 (2007), 171–184 (cit. on p. 61).

[159] Aurelio S. Tonachini. “Atmospheric monitoring with the LIDAR network of the Pierre Auger Observatory”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 617.1 (2010). 11th Pisa Meeting on Advanced Detectors, 517–519 (cit. on p. 61).

[160] V Rizi, A Tonachini, M Iarlori, G Visconti, and Pierre Auger Collaboration. “Atmospheric monitoring with LIDARs at the Pierre Auger Observatory”. *The European Physical Journal Plus* 127.8 (2012), 92 (cit. on p. 61).

[161] P. Abreu, M. Aglietta, M. Ahlers, et al. “Identifying clouds over the Pierre Auger Observatory using infrared satellite data”. *Astroparticle Physics* 50-52 (2013), 92–101 (cit. on p. 63).

[162] P. Abreu, M. Aglietta, M. Ahlers, et al. “Description of atmospheric conditions at the Pierre Auger Observatory using the Global Data Assimilation System (GDAS)”. *Astroparticle Physics* 35.9 (Apr. 2012), 591–607 (cit. on pp. 63, 163).

[163] S. Argiro, S. L. C. Barroso, J. Gonzalez, et al. “The Offline Software Framework of the Pierre Auger Observatory”. *Nucl. Instrum. Meth.* A580 (2007), 1485–1496. arXiv: 0707.1652 [astro-ph] (cit. on pp. 64, 65, 206).

[164] J. Gonzalez. “The Offline Software of the Pierre Auger Observatory:Lessons Learned”. *10th IEEE International Symposium on Parallel and Distributed Processing with Applications* (2012). arXiv: 1208.2154v1 [astro-ph.IM] (cit. on pp. 64, 65).

[165] *ROOT - Data Analysis Framework*. https://root.cern.ch (cit. on p. 64).

[166] D. Mockler. “Reconstruction of Vertical Events Recorded by the Surface Detector of the Pierre Auger Observatory”. *Proc. of the 36th International Cosmic Ray Conference, Madison, WI, USA* (2019), 353 (cit. on pp. 66, 67, 69).

[167] K. Kamata and J. Nishimura. “The Lateral and the Angular Structure Functions of Electron Showers”. *Progress of Theoretical Physics Supplement* 6 (Feb. 1958), 93–155. eprint: https://academic.oup.com/ptps/article-pdf/doi/10.1143/PTPS.6.93/5270594/6-93.pdf (cit. on p. 68).

[168] K. Greisen. *The Extenive Air Showers*. Ed. by J. G. Wilson. Vol. 3. Amsterdam North Holland Publishing Co, 1956. Chap. 1, 3–135 (cit. on p. 68).

[169] D Newton, J Knapp, and A Watson. “The optimum distance at which to determine the size of a giant air shower”. *Astroparticle Physics* 26.6 (Jan. 2007), 414–419 (cit. on p. 68).

[170] D. Mockler. “Measurement of the Cosmic Ray Spectrum with the Pierre Auger Observatory”. *PhD thesis, Universität Karlsruhe* (2018) (cit. on pp. 68, 70, 103).

[171] J. Hersil, I. Escobar, D. Scott, G. Clark, and S. Olbert. “Observations of Extensive Air Showers near the Maximum of Their Longitudinal Development”. *Phys. Rev. Lett.* 6 (1 Jan. 1961), 22–23 (cit. on p. 70).

[172] M. Unger, B.R. Dawson, R. Engel, F. Schüssler, and R. Ulrich. “Reconstruction of longitudinal profiles of ultra-high energy cosmic ray showers from fluorescence and Cherenkov light measurements”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 588.3 (Apr. 2008), 433–441 (cit. on pp. 75, 79).

[173] C. Song, Z. Cao, B.R. Dawson, et al. “Energy estimation of UHE cosmic rays using the atmospheric fluorescence technique”. *Astroparticle Physics* 14.1 (2000), 7–13 (cit. on p. 77).

[174] F. Nerling, J. Blümer, R. Engel, and M. Risse. “Universality of electron distributions in high-energy air showers—Description of Cherenkov light production”. *Astroparticle Physics* 24.6 (2006), 421–437 (cit. on p. 77).

[175] T. K. Gaisser and A. M. Hillas. “Reliability of the method of constant intensity cuts for reconstructing the average development of vertical showers”. *International Cosmic Ray Conference, 15th ICRC (Plovdiv)* 8 (1977), 353–357 (cit. on pp. 79, 206).

[176] A. Aab, P. Abreu, M. Aglietta, et al. “Data-driven estimation of the invisible energy of cosmic ray showers with the Pierre Auger Observatory”. *Phys. Rev. D* 100 (8 Oct. 2019), 082003 (cit. on p. 79).

[177] B. R. Dawson. “The Energy Scale of the Pierre Auger Observatory”. *Proc. of the 36th International Cosmic Ray Conference, Madison, WI, USA* (2019), 231 (cit. on pp. 79, 80).

[178] A. Aab, P. Abreu, M. Aglietta, et al. “Measurement of the average shape of longitudinal profiles of cosmic-ray air showers at the Pierre Auger Observatory”. *Journal of Cosmology and Astroparticle Physics* 2019.03 (Mar. 2019), 018–018 (cit. on p. 80).

[179] A. Aab et al. “The Pierre Auger Observatory Upgrade - Preliminary Design Report” (2016). arXiv: 1604.03637 [astro-ph.IM] (cit. on pp. 80–82).

[180] Antonella Castellina. “AugerPrime: the Pierre Auger Observatory Upgrade”. *EPJ Web of Conferences* 210 (2019). Ed. by I. Lhenry-Yvon, J. Biteau, O. Deligny, and P.Editors Ghia, 06002 (cit. on pp. 80–82).

[181] A. Taboada. “Analysis of Data from Surface Detector Stations of the AugerPrime Upgrade”. *Proc. of the 36th International Cosmic Ray Conference, Madison, WI, USA* (2019), 434 (cit. on pp. 80, 81).

[182] David Schmidt. “Scintillator Surface Detector simulations for AugerPrime”. *EPJ Web of Conferences* 210 (2019). Ed. by I. Lhenry-Yvon, J. Biteau, O. Deligny, and P.Editors Ghia, 06010 (cit. on p. 81).

[183] A. Aab et al. “Prototype muon detectors for the AMIGA component of the Pierre Auger Observatory”. *JINST* 11.02 (2016), P02012. arXiv: 1605.01625 [physics.ins-det] (cit. on p. 81).

[184] A. M. Botti. “The AMIGA underground muon detector of the Pierre Auger Observatory - performance and event reconstruction”. *Proc. of the 36th International Cosmic Ray Conference, Madison, WI, USA* (2019), 202 (cit. on p. 81).

[185] D. Nitz. “New Electronics for the Surface Detectors of the Pierre Auger Observatory”. *Proc. of the 36th International Cosmic Ray Conference, Madison, WI, USA* (2019), 370 (cit. on p. 82).

[186] M. Bohacova. “Test benches for the upgrade of the Pierre Auger Observatory electronics”. *Proc. of the 36th International Cosmic Ray Conference, Madison, WI, USA* (2019), 199 (cit. on p. 82).

[187] A. Aab, P. Abreu, M. Aglietta, et al. “Observation of inclined EeV air showers with the radio detector of the Pierre Auger Observatory”. *Journal of Cosmology and Astroparticle Physics* 2018.10 (Oct. 2018), 026–026 (cit. on p. 82).

[188] M. Gottowik. “Measurements of Inclined Air Showers with the Auger Engineering Radio Array at the Pierre Auger Observatory”. *Proc. of the 36th International Cosmic Ray Conference, Madison, WI, USA* (2019), 274 (cit. on p. 82).

[189] B. Pont. “A Large Radio Detector at the Pierre Auger Observatory – Measuring the Properties of Cosmic Rays up to the Highest Energies”. *Proc. of the 36th International Cosmic Ray Conference, Madison, WI, USA* (2019), 395 (cit. on p. 83).

[190] Pierre Auger Collaboration. “The Lateral Trigger Probability function for the Ultra-High Energy Cosmic Ray showers detected by the Pierre Auger Observatory”. *Astroparticle Physics* 35 (2011), 266–276. arXiv: 1111.6645v1 [astro-ph.HE] (cit. on pp. 86–88, 90, 95).

[191] M. Settimo, P. Billoir, Maris I. C., and Molina Bueno L. “Trigger probability for single stations and air-showers with the ToTd and MoPS algorithms”. *The Pierre Auger Collaboration Internal GAP Note* 2013-114 (2013) (cit. on p. 103).

[192] M. Plum. “Measurement of the chemical composition of ultra high energy cosmic rays with the HEAT telescopes of the Pierre Auger Observatory”. *PhD thesis* (2016) (cit. on pp. 110, 117, 118, 163).

[193] A. Porcelli. “Measurement of the Depth of Shower Maximum in the Transition Region between Galactic and Extragalactic Cosmic Rays with the Pierre Auger Observatory”. *PhD thesis* (2014) (cit. on pp. 111, 148, 163).

[194] Kyle Cranmer. “Kernel estimation in high-energy physics”. *Computer Physics Communications* 136.3 (May 2001), 198–207 (cit. on p. 118).

[195] Wouter Verkerke and David Kirkby. “The RooFit toolkit for data modeling” (2003). arXiv: physics/0306116 [physics.data-an] (cit. on p. 119).

[196] E. J. Ahn and others. “Measurement of the Depth of Shower Maximum of Cosmic Rays above 1 EeV”. *The Pierre Auger Collaboration Internal GAP Note* GAP2009-078 (2009) (cit. on pp. 121, 146, 149, 150).

[197] F. Meyer and F. Vernotte. “Time Tagging Board Tests at Besancon Observatory”. *The Pierre Auger Collaboration Internal GAP Note* 2001-050 (2001) (cit. on p. 123).

[198] Simon Blaess, Jose A. Bellido, and Bruce R. Dawson. “Extracting a less model dependent cosmic ray composition from X_{max} distributions” (2018). arXiv: 1803.02520 [astro-ph.HE] (cit. on p. 127).

[199] M. Unger and J. Bellido. “Supplimentary Material for the ‘long Xmax paper’”. *The Pierre Auger Collaboration Internal Note* (2014) (cit. on pp. 145, 146, 149, 150, 163–166).

[200] B. R. Dawson and others. “Proposal for an update of the Auger Energy Scale”. *The Pierre Auger Collaboration Internal GAP Note* 2012-124 (2012) (cit. on p. 146).

[201] S. Falk, R. Engel, R. Ulrich, and M. Unger. “Telescope Alignment Studies”. *The Pierre Auger Collaboration Internal GAP Note* 2011-123 (2011) (cit. on p. 148).

[202] C. De Donato and others. “Using star tracks to determine the absolute pointing of the Fluorescence Detector telescopes”. *The Pierre Auger Collaboration Internal GAP Note* 2005-008 (2005) (cit. on p. 148).

[203] M. Prouza and others. “Star tracking using background data of FD telescopes”. *The Pierre Auger Collaboration Internal GAP Note* 2005-041 (2005) (cit. on p. 148).

[204] M. Scuderi and others. “Alignment method of the Pierre Auger Observatory fluorescence telescopes based on the Central Laser Facility events”. *The Pierre Auger Collaboration Internal GAP Note* 2006-061 (2006) (cit. on p. 148).

[205] P. Younk and B. Fick. “FD Telescope Alignment Using CLF Laser Shots”. *The Pierre Auger Collaboration Internal GAP Note* 2006-087 (2006) (cit. on p. 148).

[206] J. Bellido. “Tracking HEAT and CO calibrations and the effects on Xmax for the HeCo system”. *The Pierre Auger Collaboration Analysis Meeting* (2017) (cit. on pp. 150–152).

[207] M D Roberts. “The role of atmospheric multiple scattering in the transmission of fluorescence light from extensive air showers”. *Journal of Physics G: Nuclear and Particle Physics* 31.11 (Sept. 2005), 1291–1301 (cit. on p. 164).

[208] J. Pekala, P. Homola, B. Wilczyńska, and H. Wilczyński. “Atmospheric multiple scattering of fluorescence and Cherenkov light emitted by extensive air showers”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 605.3 (2009), 388–398 (cit. on p. 164).

[209] M. Unger, R. Engel, F. Schussler, and R. Ulrich. “Lateral shower light distributions in the Cherenkov-Fluorescence-Matrix Profile Reconstruction”. *The Pierre Auger Collaboration Internal GAP Note* 2008-052 (2008) (cit. on p. 164).

[210] K. Louedec and J. Colombi. “Effect of the aerosol size on multiple scattering: Proposal for an updated systematic uncertainty on energy and depth of maximum”. *The Pierre Auger Collaboration Internal GAP Note* 2013-116 (2013) (cit. on p. 164).

[211] S. Knurenko and I. Petrov. “Mass composition of cosmic rays above 0.1 EeV by the Yakutsk array data”. *Advances in Space Research* 64.12 (Dec. 2019), 2570–2577 (cit. on p. 182).

[212] V.V. Prosin, S.F. Berezhnev, N.M. Budnev, et al. “Primary CR energy spectrum and mass composition by the data of Tunka-133 array”. *EPJ Web of Conferences* 99 (2015), 04002 (cit. on p. 182).

[213] S. Buitink, A. Corstanje, H. Falcke, et al. “A large light-mass component of cosmic rays at 1017–1017.5 electronvolts from radio observations”. *Nature* 531.7592 (Mar. 2016), 70–73 (cit. on p. 182).

[214] R. U. Abbasi, M. Abe, T. Abu-Zayyad, et al. “Depth of Ultra High Energy Cosmic Ray Induced Air Shower Maxima Measured by the Telescope Array Black Rock and Long Ridge FADC Fluorescence Detectors and Surface Array in Hybrid Mode”. *The Astrophysical Journal* 858.2 (May 2018), 76 (cit. on p. 182).

[215] T. Fujii et al. “Detection of ultra-high energy cosmic ray showers with a single-pixel fluorescence telescope”. *Astropart. Phys.* 74 (2016), 64–72. arXiv: 1504.00692 [astro-ph.IM] (cit. on pp. 187, 188, 196–198, 273–275).

[216] T. Fujii et al. “A next-generation ground array for the detection of ultrahigh-energy cosmic rays: the Fluorescence detector Array of Single-pixel Telescopes (FAST)”. *EPJ Web Conf.* 210.06003 (2019) (cit. on p. 187).

[217] M. Malacari, J. Farmer, T. Fujii, et al. “The first full-scale prototypes of the fluorescence detector array of single-pixel telescopes”. *Astroparticle Physics* 119 (2020), 102430 (cit. on pp. 187, 189–195, 202, 205, 207, 229, 297).

[218] D. Mandat et al. “The prototype opto-mechanical system for the Fluorescence detector Array of Single-pixel Telescopes”. *JINST* 12.07 (2017), T07001 (cit. on pp. 187, 189).

[219] P. Sommers. “Capabilities of a giant hybrid air shower detector”. *Astroparticle Physics* 3 (1995), 349–360 (cit. on p. 187).

[220] T. Fujii et al. “The FAST Project - A Next Generation UHECR Observatory”. *EPJ Web Conf.* 136.02015 (2017) (cit. on p. 188).

[221] T. Fujii et al. “The Full-Scale Prototype for the Fluorescence Detector Array of Single-Pixel Telescopes”. *JPS Conf. Proc.* 19.011047 (2018) (cit. on pp. 188, 191).

[222] D. Mandat, M. Palatka, M. Pech, P. Schovanek, P. Travnicek, et al. “The Prototype Opto-mechanical System for the Fluorescence detector Array of Single-pixel Telescopes”. *PoS* ICRC2017 (2017), 389 (cit. on p. 189).

[223] M. Pech et al. “Simulation of the optical performance of the Fluorescence detector Array of Single-pixel Telescopes”. *EPJ Web Conf.* 210.05014 (2019) (cit. on pp. 189, 190).

[224] R. Abbasi, Y. Abdou, T. Abu-Zayyad, et al. “Calibration and characterization of the IceCube photomultiplier tube”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 618.1 (2010), 139–152 (cit. on p. 189).

[225] D. Mandat et al. “Atmospheric monitoring with the Fluorescence detector Array of Single-pixel Telescopes”. *EPJ Web Conf.* 197 (2019), 02002 (cit. on p. 193).

[226] F. Schussler. “Top-Down Reconstruction of Ultrahigh Energy Air Showers measured with the Pierre Auger Fluorescence Detector”. *Diploma thesis, Universität Karlsruhe* (2005) (cit. on p. 200).

[227] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016 (cit. on pp. 227, 236).

[228] F. Chollet and others. “Keras” (2015) (cit. on p. 237).

[229] Martin Abadi, Ashish Agarwal, Paul Barham, et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015 (cit. on p. 237).

[230] Diederik P. Kingma and Jimmy Ba. “Adam: A Method for Stochastic Optimization” (2014). arXiv: 1412.6980 [cs.LG] (cit. on p. 237).

[231] Dusan Mandat (2020). Private Communication (cit. on p. 253).

[232] B. Keilhauer, R. Engel, B. Wilczynska, H. Wilczynski, and M. Will. “Lateral shower light distributions in the Cherenkov-Fluorescence-Matrix Profile Reconstruction”. *The Pierre Auger Collaboration Internal GAP Note* 2009-037 (2009) (cit. on p. 257).

Colophon

This thesis was typeset with LaTeX 2_ε. It uses the *Clean Thesis* style developed by Ricardo Langner. The design of the *Clean Thesis* style is inspired by user guide documents from Apple Inc.

Download the *Clean Thesis* style at http://cleanthesis.der-ric.de/.